

OM of: US-09-303-518D-463 to: SPTREMBL\_19:★ out\_format : pfs  
Date: Jun 30, 2002 8:07 AM

Command line parameters:

```
-MODEL-frame-nzp.model -DEV-xlh
-o/cqnr2.1/USPTO.spool/US09303518/runat.28062002.142714.4338/app_query.fasta.1.235051
-BD=SPRMBLH.19 -QEMT=fastan -SUFFIX=1TSP -GAPO=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GCAPOP=4.500 -GCAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCAPOP=6.000 -FCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXF=7.000 -START=1 -MATRIX=blonsurf62
-TRANS=human40.cdi -LIST=100 -DOCALLIGN=200 -THR SCORE=spct
-THR MAX=100 -THR MIN=0 -ALIGN=45 -MODE=LOCAL -OUTPMT=pf5
-NORM-ext -HEAPSIZ=500 -MINTEN=0 -MAXLEN=2000000000
-USER=US09303518 @GCNL.1.1967 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XIPXY -WAIT -THREADS=1
```

Search information block:

Query: US-09-303-518D-463

Query Length: 1887

Database: SPTREMBL\_19:\*

Database sequences: 56222 Database length: 17300403

Database length: 1/2994929  
Search time (sec): 883.700000

Sequence	Strd Orig	ZScore	EScore	Len	Document
sp_bacteriap:09JUG1	+ 2200.00	2594.46	6.5e-137	556	09JUG1 neisseria meningitidis
sp_bacteriap:09KVF3	+ 1016.50	1136.95	6.1e-59	412	09KVF3 neisseria meningitidis
sp_bacteriap:09K120	+ 752.50	883.30	1.6e-41	467	09K120 neisseria meningitidis
sp_bacteriap:09JUSV9	+ 740.50	868.45	1.0e-40	498	09JUSV9 neisseria meningitidis
sp_bacteriap:09KOF4	+ 658.00	772.48	2.6e-35	422	09KOF4 neisseria meningitidis
sp_bacteriap:09JUKF8	+ 450.00	524.47	1.4e-21	459	09JUKF8 neisseria gonorrhoeae
sp_bacteriap:09JUKR8	+ 414.50	483.01	3.1e-19	582	09JUKR8 neisseria meningitidis
sp_plant:09FYB2	+ 390.00	453.99	1.3e-17	483	09FYB2 arabidopsis thaliana (m
sp_plant:09ZMX2	+ 172.00	189.88	0.0035	894	09ZMX2 arabidopsis thaliana (m
sp_plant:09ZM08	+ 163.50	179.86	0.0127	891	09ZM08 arabidopsis thaliana (m
sp_invertebrate:09NMS7	+ 162.50	176.57	0.0153	1079	09NMS7 caenorhabditis elegans
sp_invertebrate:020908	+ 162.00	179.54	0.0136	770	020908 caenorhabditis elegans
sp_bacteriap:09J9N9E2	+ 161.50	182.69	0.0149	529	09J9N9 microbacterium ammoni
sp_bacteriap:09J9NE2	+ 158.50	194.27	0.0154	116	09J9NE2 pseudomonas fluoresc
sp_human:09J9SE52	+ 158.00	159.51	0.0432	3570	09J9SE52 homo sapiens (human)
sp_bacteriap:053478	+ 157.50	179.44	0.0262	456	053478 neisseria gonorrhoeae
sp_human:09J0039	+ 157.00	171.47	0.0348	936	09J0039 homo sapiens (human)
sp_human:09J0U8	+ 157.00	168.70	0.0376	1262	09J0U8 homo sapiens (human)
sp_human:09J0035	+ 157.00	162.73	0.0444	2296	09J0035 homo sapiens (human)
sp_human:09J0P82	+ 155.50	160.92	0.0467	2752	09J0P82 homo sapiens (human)
sp_human:076037	+ 155.50	169.82	0.0436	944	076037 homo sapiens (human)
sp_human:09J4MD6	+ 155.50	168.80	0.0448	1045	09J4MD6 homo sapiens (human)
sp_human:09J6ZM2	+ 155.50	168.23	0.0455	1107	09J6ZM2 homo sapiens (human)
sp_human:09J6V6	+ 155.50	167.79	0.0461	1157	09J6V6 homo sapiens (human)
sp_human:09J4H87	+ 155.50	167.53	0.0466	1187	09J4H87 homo sapiens (human)
sp_human:09J4H85	+ 155.50	167.41	0.0466	1199	09J4H85 homo sapiens (human)
sp_human:09J4H84	+ 155.50	167.31	0.0467	1214	09J4H84 homo sapiens (human)
sp_human:09J4H83	+ 155.50	167.30	0.0467	1215	09J4H83 homo sapiens (human)
sp_human:09J4H82	+ 155.50	166.97	0.0472	1256	09J4H82 homo sapiens (human)
sp_human:09J4H88	+ 155.50	163.33	0.0524	1827	09J4H88 homo sapiens (human)
sp_human:09J4H81	+ 155.50	161.76	0.0546	2117	09J4H81 homo sapiens (human)
sp_human:09J4H80	+ 155.50	161.53	0.0549	2167	09J4H80 homo sapiens (human)
sp_human:014681	+ 154.00	161.52	0.0549	2159	014681 homo sapiens (human)
sp_human:09J4H89	+ 154.00	172.20	0.0487	622	09J4H89 homo sapiens (human)
sp_fungi:09J22495	+ 152.00	165.41	0.0744	897	09J22495 emeticella ridulans (sag
sp_rickettsia:070495	+ 150.50	164.41	0.0917	961	070495 mus musculus (mouse)
sp_human:09J0U36	+ 150.00	160.31	0.1091	1275	09J0U36 homo sapiens (human)
sp_human:015038	+ 150.00	156.97	0.1198	1783	015038 homo sapiens (human)
sp_human:060382	+ 150.00	156.92	0.1200	1791	060382 homo sapiens (human)

seq_name: sp_bacteriap:Q9JVG1	seq_documentation_block: PRELIMINARY;	PRT;	556 AA.				
ID Q9JVG1							
AC Q9JVG1							
DT 01-OCT-2000 (TREMblrel. 15, Created)							
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)							
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)							
DE PUTATIVE ADHESIN.							
GN MARB3 OR NMA0853.							
NC Neisseria meningitidis (serogroup A).							
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.							
OX NCBI_TaxId:65699;							
sp_human:Q93418	+	149.00	168.21	0.0985	513	! Q43418	homo sapiens (human)
sp_human:Q9N261	+	148.00	153.70	0.1258	716	! Q9N44	homo sapiens (human)
sp_bacteria:Q9Z4N7	+	148.00	154.11	0.1646	1873	! Q9Z47	enterococcus faecalis (human)
sp_mammal:Q23071	+	145.50	163.78	0.1689	528	! Q23071	sus scrofa (pig)
sp_invertebrate:Q9VU43	+	145.00	157.29	0.2150	954	! Q9V43	drosophila melanogaster (fruit fly)
sp_rodent:Q993P2	+	145.00	149.78	0.2652	2025	! Q993P2	mus musculus (mouse)
sp_bacteria:Q985O18	+	144.00	154.48	0.2618	1122	! Q985O18	mycobacterium smegmatis (bacterium)
sp_rodent:Q95361	+	143.50	149.15	0.3225	1805	! Q95361	rattus norvegicus (rat)
sp_human:Q9N5M8	+	142.00	150.66	0.3694	1299	! Q9N5M8	homo sapiens (human)
sp_invertebrate:Q76894	+	142.00	147.43	0.4043	1795	! Q76894	drosophila melanogaster (fruit fly)
sp_invertebrate:Q9VXMS	+	142.00	144.51	0.4387	2406	! Q9VXMS	drosophila melanogaster (fruit fly)
sp_rodent:Q70670	+	141.50	144.87	0.4609	2187	! P70670	mus musculus (mouse)
sp_human:Q14879	+	141.00	158.17	0.3371	543	! Q14879	homo sapiens (human)
sp_human:Q95085	+	141.00	154.06	0.3782	830	! Q95085	homo sapiens (human)
sp_human:Q9H306	+	140.50	152.80	0.4157	877	! Q9H306	homo sapiens (human)
sp_human:Q14651	+	140.50	152.79	0.4158	878	! Q9H346	homo sapiens (human)
sp_human:Q90UKW9	+	140.50	151.93	0.4259	957	! Q9H37	homo sapiens (human)
sp_human:Q9H015	+	139.50	149.53	0.4555	1217	! Q9UKW9	homo sapiens (human)
sp_human:Q9C012	+	139.00	150.40	0.4555	991	! Q9H15	homo sapiens (human)
sp_bacteriap:Q9CNC9	+	138.00	170.64	0.3324	1236	! Q9C12	homo sapiens (human)
sp_human:Q9N5N9	+	138.00	145.82	0.6798	109	! Q9N5N9	pasturella multocida (bacterium)
sp_rodent:Q920R4	+	138.00	143.16	0.7344	1312	! Q9N35	homo sapiens (human)
sp_invertebrate:Q95Z27	+	137.50	146.85	0.7010	1714	! Q9Z074	mus musculus (mouse)
sp_bacteria:Q931B1	+	137.00	159.39	0.5237	1116	! Q931B1	caenorhabditis elegans (nematode)
sp_invertebrate:Q10930	+	137.00	149.01	0.7002	299	! Q931B1	bifidobacterium animalis (bacterium)
sp_invertebrate:Q994U5	+	136.50	161.07	0.5301	847	! Q10930	caenorhabditis elegans (nematode)
sp_invertebrate:Q960E9	+	136.50	152.29	0.6777	574	! Q960E9	drosophila melanogaster (fruit fly)
sp_invertebrate:Q9V515	+	136.50	149.68	0.7291	746	! Q9V515	drosophila melanogaster (fruit fly)
sp_virus:Q90P87	+	136.50	147.96	0.7650	886	! Q90P87	human herpesvirus 4 (virus)
sp_human:Q92954	+	136.50	143.37	0.8698	1404	! Q92954	homo sapiens (human)
sp_virus:Q956854	+	136.00	147.72	0.8174	856	! Q956854	human herpesvirus 4 (virus)
sp_fungi:Q929370	+	136.00	136.67	1.11	2592	! Q929370	neurospora crassa (fungus)
sp_plant:Q230354	+	135.00	154.07	0.7705	402	! Q230354	arabidopsis thaliana (plant)
sp_fungi:Q07229	+	135.00	147.00	0.9381	817	! Q07229	saccharomyces cerevisiae (yeast)
sp_invertebrate:Q44548	+	135.00	146.29	0.9378	877	! Q44548	caenorhabditis elegans (nematode)
sp_human:Q947N4	+	135.00	143.04	1.08	1343	! Q947N4	homo sapiens (human)
sp_rodent:Q984491	+	135.00	135.50	1.30	2588	! Q984491	mus musculus (mouse)
sp_invertebrate:Q9WZ21	+	134.50	140.49	1.20	1479	! Q9WZ21	drosophila melanogaster (fruit fly)
sp_invertebrate:Q22579	+	134.50	138.29	1.27	1844	! Q22579	caenorhabditis elegans (nematode)
sp_invertebrate:Q117921	+	134.00	146.42	1.07	769	! Q117921	caenorhabditis elegans (nematode)
sp_bacteria:Q47692	+	134.00	140.61	1.26	1377	! Q47692	escherichia coli (bacterium)
sp_fungi:Q42854	+	134.00	143.04	1.32	1611	! Q42854	schizosaccharomyces pombe (yeast)
sp_plant:Q9P038	+	133.50	143.07	1.25	1014	! Q9P038	oryza sativa (rice)
sp_invertebrate:Q57580	+	133.50	141.81	1.30	1151	! Q57580	gallus gallus (chicken)
sp_bacteriap:Q9A6A1	+	133.00	151.17	1.06	424	! Q9A6A1	caldococcus thermophilus (bacterium)
sp_fungi:Q87107	+	133.00	144.25	1.29	849	! P87107	saclibacter crescentus (bacterium)
sp_human:Q99621	+	133.00	140.88	1.41	1190	! Q99621	saccharomyces cerevisiae (yeast)
sp_invertebrate:Q9BKV7	+	133.00	139.81	1.46	1125	! Q9BKV7	leishmania major (protozoan)
sp_human:Q9Y474	+	133.00	137.59	1.55	1656	! Q9Y474	homo sapiens (human)
sp_invertebrate:Q9W204	+	132.50	140.38	1.52	1179	! Q9W204	drosophila melanogaster (fruit fly)
sp_human:Q94945	+	132.00	139.77	1.64	1180	! Q94945	homo sapiens (human)
sp_virus:Q9NMH9	+	132.00	139.29	1.66	1242	! Q9NMH9	rat cytomegalovirus (virus)
sp_invertebrate:Q9W0B0	+	132.00	137.87	1.73	1430	! Q9W0B0	drosophila melanogaster (fruit fly)
sp_archaea:Q9H456	+	131.50	144.48	1.53	694	! Q9H456	halobacterium sp. (st
sp_human:Q99X45	+	131.50	137.46	1.86	1404	! Q99X45	homo sapiens (human)
sp_human:Q99116	+	131.50	135.82	1.94	1654	! Q99116	homo sapiens (human)
sp_mammal:Q18984	+	131.00	147.99	1.47	460	! Q18984	cercopithecus aethiops (primate)



480 rgtYrIleProProLysAsnTYrSerProSerAlaProLeuProLysGly 496  
1443 ..... 1443  
497 ProAsnAsnGlyTYrLeuAspLysPheGlyAsnGluTYrThrLysGlyPr 513  
1444 .AGAAGAGTACAGTACAGTACAGTAAAGCCATCGCCAA..... 1482  
513 oSerArgThrLysGlyGlnGluPheGluTYrPaspAlaGlnLeuSerLysT 530  
1483 .....CGAGA.....TGGCAATAAAGAGGTTAGATTTAAT 1518  
530 hGlyArgGlnGlnLeuGlnTYrPalaSerArgAspGlyHisLeuAsn 546  
1519 CATTTATAGTGTGATATCAATAGAA 1548  
547 lIeSerIleAspGlyLysIleThrHisLys 556

seq\_name: sp\_bacteriap:09KOF9

seq\_documentation\_block:

ID 09KOF9 PRELIMINARY; PRT; 412 AA.  
AC 09KOF9;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)  
DE HYPOTHETICAL PROTEIN NMB0647.  
GN NMB0647.  
OS Neisseria meningitidis (serogroup B).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI:taxid:491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MC58 / SEROGROUP B;  
RX MEDLINE-20175755; PubMed-10710307;  
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
RA Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K.,  
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
RA Mason T., Clecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,  
RA Cotton M.D., Ueberback T.R., Khouri H., Qin H., Vamathevan J.,  
RA Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,  
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
MC58.";  
RL Science 287:1809-1815(2000).  
DR EMBL; AEO02419; AAF41068.1; -;  
DR TIGR; NMB0647; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 412 AA; 45508 MW; 6B41BAD5323C9052 CRC64;

alignment\_scores:

Quality: 1016.50 Length: 449  
Ratio: 3.592 Gaps: 12  
Percent Similarity: 63.029 Percent Identity: 51.448

alignment\_block:

US-09-303-518D-463 x 09KOF9

Align seg 1/1 to: 09KOF9 from: 1 to: 412

588 GAGCAAGAGATGAGCGAGGATTT.....CAAGCGG 619  
||| ..... |||  
42 AspLysGlnThrAlaArgLysIleTYrSerValIleLysGlnGlnValVal 58  
620 CCACCCGATACAGCCCGAGCTGGACAGATCGGCAATGCCCGCGAAGC 669  
||| ..... |||  
58 shisGluCysSer...SerThrGlyArgIleThrGluCysArgGlnAsnI 74  
670 TTCAAGCGGACATGAGATATCTCAAAACATCATCGCGCGGACAGAGAG 719  
:: ||| ..... |||  
74 legIaArgIleIleGluPheThrGlnAsp..... 83

720 AATTGCGCGCAGAGCGATGCCGTGACAGGTATTAAGCAAGCTCAACA 769  
::: ||| ..... |||  
84 .....LysArgPheAspSerArgPheLysAspLeuLysGly 96  
770 TTGCTGATGACACAGCGCTTGCTGCTTCCACCGAACAAGATGGCG 819  
: ..... |||  
96 u.....SerLeuTYrTYrLeuAsnLysHisPro 105  
820 CGCATCAACGATTTGGCAGATATGGCGCAACCTCAAGACTATGCCGAGC 869  
||| ||| ..... |||  
106 .....AspLeuValAlaSerTYrLeuLysAlaGluTYrGluLysLe 119  
870 AGCCATCCGCGATTTGGCGATCCCAAAACCCCAATGGCGCAAGCATAG 919  
::: ||| ..... |||  
119 uAspArgGluAspLysSerIleLeuHisArgTYrIleSerProGlyAlaG 136  
920 AAGCCGTCAGC...AATATCTTATGGCAGCCATCCCATCAAGGAT 966  
|| ||| ||| ..... |||  
136 lIuIleValSerGlySerLeuGlyValValLeuSerGlyValAlaGlyGly 152  
967 GAGCTGTCCGGGAAATACGGCTTGGC.....GG 998  
||| ..... |||  
153 GlySerCysAlaGluThrPheGlyLeuGlyCysAlaAlaLeuValGlu 169  
999 CATCAGCGCA.....CATCTGTCAAGCGGTGCGCAGATGGCGCGA 1039  
||| ||| ..... |||  
169 yAlaThrSerSerTYrAspHisValIleThrGlyThrLys..... 182  
1040 TCCGATTCGCGAAGGAATCCGCCGTGACGACAAATTTGCCGATGG 1089  
183 ..... ||| |||  
1090 GCATACGCCAATATACCCGTCCCTTACATCCCGAATATCGCTCAAA 1139  
||| ..... |||  
188 AlaSerGlnGlnArgPro..... 193  
1140 CTGGAGACAGCGTTAGCGCAAGAAACATCACTCTCTCAACCGTCGCC 1189  
194 ..... ThrIleAlaVal 197  
1190 CGTCAACGCCAATAATGTCAAACTGGACCAACGCCACCGCAAGACA 1239  
::: ||| ..... |||  
197 alGlnAlaLeuLysGlnLeuGlyLeuSerGlnGlnAla..... 209  
1240 GCGGTACGTTTGACGCTAAAGGTTCCGAATTTGAGACACAGCTGAA 1289  
210 ..... AlaGluTYrValGlu 214  
1290 ATATGATACGAAGCTCGATATTCAGAAATTATCGGGGGCGGTATACCTA 1339  
::: ||| ..... |||  
214 nPheSerIleAspLeuPheSerValGlyLysSerGlyGlyIleProL 231  
1340 AGGCTAAGCCTGTGTTGATGCGAAACCGAGATGGAGGTGATAGAG 1389  
||| ..... |||  
231 yAlaLysProValPheAspAlaLysProArgTYrGluValAspArgLys 247  
1390 CTTAATTAATTTACAACTCGTGCAGAGGTGGGAAATTTTCAGAGAAC 1439  
||| ..... |||  
248 LeuAsnLysLeuThrThrArgGlnGlnValGlnLysAsnValGlnGluTh 264  
1440 GAGAGAGAGAGTACAGTACAGTACAGTAAAGCCATCGGCAAGAGAT 1489  
||| ..... |||  
264 rArgArgArgSerGlnSerSerGlnPheLysAlaHisAlaGlnArgGlu 281  
1490 GGGAAATTAACAGAGGTAGATTTTAATCATTTTATAGTGGTGTATTC 1539  
||| ..... |||  
281 rGluAsnLysThrGlyLeuAspPheAsnHisPheIleGlyLysPhe 297  
1540 AATAGAAAGCAGACAGTAAACAGAGGCGATAGTCAACCGGTGTATGT 1589  
||| ..... |||  
298 AsnLysLysGlyTYrValThrGlyGlyHisSerLeuThrArgLysAsp 314

1590 ACGGGGATACCAACAACACTGGCACCGCATAAACATGGGTTATACAG 1639  
314 IarYvalliEgIngInThSerAlaProaspIyHisGlyValIyrgIna 331  
1640 CACACAGTGAATTAATAAAGCCTGATGAGTGGAGTGAATAAACGAA 1689  
331 IarThValGluIleIySlySProaspGlySerTrpGluValIyShrIyS 347  
1690 AAGGTGGCAAGTATGATGACCAACACACACATGTTCCCAAAAGATTGGGA 1739  
348 LysGlyGlyLysValMetThrIyShIsthMetPheProIySAspTrpAs 364  
1740 TGAAGCTAGATTAAGCGCTGAAAGTACTTCGGCTTGGCAAAAGTACAAATA 1789  
364 pGluAlaArgIleArgAlaGluValThSerIaTrpGluSerArgIleM 381  
1790 TGCTTAGAGTATATAATGACGAGGGTACACAGTAATTCGGCTTTAAATA 1839  
381 eIeIuIySAspSnuIyStrpGInGlyThSerLysSerGlyIleIyS 397  
1840 GAGGATTTACCAACCTAATGAGACAGCATATCCCATTTATGAA 1884  
398 GluGlyPheThrGluProaspIyAlaArgIleAlaArgIleProIleIyGlu 412

```

320 ..... 320
1001 TCACGCACATCTCTCAAGGGGTCGAGATGGCGCGATTCGCG 1050
321 ..... LysValAlaLysLeuAlaLysAlaLysPro 331
1051 AAGGGAATCCGCCGTCAGCAATTTGCCGATCGGCATCGCCAA 1100
332 ...GlyLysAlaAlaValSerGlyAspPheAlaAspSerTyrLysLys 347
1101 ATACCCGCCCTTACCATTCGCCAAATATCCGTTCAAACTTGAGCAGC 1150
347 sLeuAlaLeuSerAspSerAlaArgInLeuTyrGlnAsnAlaLysTyr 364
1151 GTACGCGAAGAAACATCATCTCTCAACCCGCGCGTCAAGGC 1200
364 rg.....GluAla 366
1201 AAAATGTCAAACTGGCAGACCAACGCCCGAAGACAGCGCTACCGTT 1250
367 LeuAspIleHisTyrGlnAspLeuIleArgArgLysThr..... 379
1251 TGACGGTAAGGGTTTCCGAATTTTGAAAGCACAGCTAAATATGATGCA 1300
380 .AspGlySer.....LysPheIleAsnG 388
1301 AGCTCATATTTCAGAAATTAACGGGGCGGTATACCTAGGCTAAGCCT 1350
388 LysArgGlnIleAspAlaValThrAsnAspAlaLeuIleGlnAlaLysAsn 404
1351 GTGTTATATGCGAAACCGAGATGGAGTTGATAGAACTTAATAATT 1400
405 ThrIleSerAla.....IleAspLysProLysAsnPhe 416
1401 GACA.....ACTGCTAGCAGGAGGAGAAATTTTCAGAAACAGAA 1444
416 uAsnGlnLysAsnArgLysGlnIleLysAlaThrIleGlnAlaLysAsn 433
1445 GAAAGACTCAGAGTACTGCTTT.....AAGCCCATGGCGAA 1482
433 InGlnGlyLysArgAlaGlnPheThrPheLysTyrGlyAlaHisSerGln 449
1483 CGAGAA.....TGGCAAAATTAACAGGCTTA 1509
450 ValLysSerTyrIleGlnSerLysGlyGlyIle 460

seq_name: sp_bacteriap:09JSV9
seq_documentation_block:
ID 09JSV9 PRELIMINARY; PRT; 498 AA.
AC 09JSV9;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ADHESIN MAFB2
GN MAFB2 OR NMA2113
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20226556; PubMed=10761919;
RA Parikh H.J., Achman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holtroyd S.,
RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491.
RL Nature 404:502-506(2000).
DR EMBL; AL162758; CAB5327.1; -

```

```

KW Complete proteome.
SQ SEQUENCE 498 AA; 54653 MW; 38F1E7C12B8811F7 CRC64;

alignment_scores:
Quality: 740.50 Length: 586
Ratio: 2.165 Gaps: 18
Percent Similarity: 58.362 Percent Identity: 32.253

alignment_block:
US-09-303-518D-463 x 09JSV9
Align seg 1/1 to: 09JSV9 from: 1 to: 498

28 ATTGTCATCACTGGCAGTGTG.....CTGCCGAT 59
3.....
7 LeuIleLysLeuLeuAlaLysAlaLysAlaValAlaAlaLysLeuIleG 23
60 GCATGCACAGCCCTCAGATTGGCAACAGATCCCTTTTCCGCGAGCTTC 109
23 nProAlaLeuAlaAlaAspLeuAlaGlnAspProPheIleThrAspAsn 40
110 TCGACCGTCAGCATTTGCAACCCGACGGAATACCACTATTGGCAGC 159
40 LagIlaArgGlnHisTyrGlnProGlyGlyLysTyrHisLeuPheGlyAsp 56
160 ...AGGGGGAGCTTGCNAGCGCAAGCGCATATCGGATTGGAAACAT 206
57 ProArgGlySerValSerAspArgThrGlyGlnIleAsnValIleGlnAs 73
207 ACAAGCCATCAGTTGGGCGCCTGATGATGATCAACAGCGCGCGCTGAAG 256
73 pTyrThrHisArgMetGlyLeuLeuIleGlnIleAsnIleAsnG 90
257 GAAATATCGGTACTATGCTTCGCTTTTCCGATCAGCGGACAAATTCAT 306
90 LysThrIleGlyLysHisThrArgPheSerGlyHisGlyTyrGlnIleHis 106
307 TCGCCCTTCGACACCATAGCCATTCGATTCGATTCGACGAAACCGGTAG 356
107 AlaProPheAspAsnHisAlaAlaAspSerAlaSerGlnGlyLys 123
357 TCCCGTTGACGATTCAGCCTTTACCGATTCGATTCAGGAGGATCGAGAC 406
123 nValAspGlnGlyPheThrValTyrArgLeuAsnTrpIleGlnHisGln 140
407 ACCATTCGCGCGACGGCTATGACGGGCCACAGGGCGCGCTATCCGCT 456
140 IsnHisProAlaAspAlaTyrAspGlyProLysGlyLysTyrProLys 156
457 CCCAAAGCGCGAGGATATATACAGCTAGCAGCATAAAGCGCTTGCCCA 506
157 ProThrGlyAlaArgAspGlyTyrThrTyrHisValAsnGlyThrAlaAr 173
507 AAATATCCGCTCAACCTGACGACCAACCGCAACCGGACCGGAGCGTGG 556
173 gSerIleLysLeuAsnProThrAspThrArgSerIleArgInArgIles 190
557 CGACCGCTTTCACATGCCCGCGCTATGCTGACGCAAGAGTARGGCGAC 606
190 eAspAsnTyrSerAsnLeuGlySerAsnPheSerAspAlaAspGln 206
607 GGATTCAACAGCGCCGACCATACAGCCCGGAGCTGACAGATCGGGCAA 656
207 AlaAsnArgLysMetPheGlnHisAsnAlaLysLeuAspArgTyrGlyAs 223
657 TGCCGCGAAGCCTTCAACGACCATCGACATATGCTCAAAACATCATCG 706
223 nSerMetGlnPheIleAsnGlyValAlaAlaGlyAlaLeuAsnProHe 240
707 GCGCGGAGGAGAAATGTCGCGCGACGATGCGCTGACGAGATTAAGC 756
240 LeSerAlaGlyAlaLeuGlyIleGlyAspIleLeuTyrGlyThrArg 256

```

757	GAAGGCTCAAAACATTGGCTGCATCCACAGCGCTGGCTGCTTCACACGA	806
758	.....	807
759	.....	808
760	.....	809
761	.....	810
762	.....	811
763	.....	812
764	.....	813
765	.....	814
766	.....	815
767	.....	816
768	.....	817
769	.....	818
770	.....	819
771	.....	820
772	.....	821
773	.....	822
774	.....	823
775	.....	824
776	.....	825
777	.....	826
778	.....	827
779	.....	828
780	.....	829
781	.....	830
782	.....	831
783	.....	832
784	.....	833
785	.....	834
786	.....	835
787	.....	836
788	.....	837
789	.....	838
790	.....	839
791	.....	840
792	.....	841
793	.....	842
794	.....	843
795	.....	844
796	.....	845
797	.....	846
798	.....	847
799	.....	848
800	.....	849
801	.....	850
802	.....	851
803	.....	852
804	.....	853
805	.....	854
806	.....	855
807	.....	856
808	.....	857
809	.....	858
810	.....	859
811	.....	860
812	.....	861
813	.....	862
814	.....	863
815	.....	864
816	.....	865
817	.....	866
818	.....	867
819	.....	868
820	.....	869
821	.....	870
822	.....	871
823	.....	872
824	.....	873
825	.....	874
826	.....	875
827	.....	876
828	.....	877
829	.....	878
830	.....	879
831	.....	880
832	.....	881
833	.....	882
834	.....	883
835	.....	884
836	.....	885
837	.....	886
838	.....	887
839	.....	888
840	.....	889
841	.....	890
842	.....	891
843	.....	892
844	.....	893
845	.....	894
846	.....	895
847	.....	896
848	.....	897
849	.....	898
850	.....	899
851	.....	900
852	.....	901
853	.....	902
854	.....	903
855	.....	904
856	.....	905
857	.....	906
858	.....	907
859	.....	908
860	.....	909
861	.....	910
862	.....	911
863	.....	912
864	.....	913
865	.....	914
866	.....	915
867	.....	916
868	.....	917
869	.....	918
870	.....	919
871	.....	920
872	.....	921
873	.....	922
874	.....	923
875	.....	924
876	.....	925
877	.....	926
878	.....	927
879	.....	928
880	.....	929
881	.....	930
882	.....	931
883	.....	932

```

1621 AAACATGGGGCTTATCAACGCACAGTGGAAATTAATAAAGCGCTGATGGAG 1670
      :: |||:: |||||
465 ArgaspGlyIleArgTleArgThrVal..... 473
1671 TTGGGAGCGTGAACGCAAAAAAGTGGGAAAGTGATGACCAACGACACCA 1720
      ::|||:: ||| |||||::||| |||
474 TYRGUProAlaThr.....GlyIysValVal.....Thra 484
1721 TCGTCCCA 1728
      |||||
484 laphnePro 486

seq_name: sp_bacteriap:Q9K0F4

seq_documentation_block:
ID   Q9K0F4   PRELIMINARY;   PRT;   422 AA.

```

DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE MAFB-RELATED PROTEIN.  
GN NMB00653.  
OS *Neisseria meningitidis* (serogroup B).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; *Neisseria*.  
OX NCBI\_TaxID:491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MC58 / SEROGROUP B;  
RX MEDLINE-2017555; PubMed-10710307;  
RA Tetselin H., Saunders N.J., Heldelberg J., Jeffries A.C., Nelson K.E.,  
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
RA Nelson W.C., Gwyn M.L., Deboy R., Peterson J.D., Hickey E.K.,  
RA Hatt D.H., Salberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
RA Mason T., Clecko A., Parksey D.S., Blair E., Cillone H., Clark E.B.,  
RA Cotton M.D., Ulfersbach T.R., Khouri H., Qin H., Vamathevan J.,  
RA Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,  
RA Smith H.O., Fraser C.M., Moran E.R., Rappunli R., Venter J.C.;  
RT "Complete genome sequence of *Neisseria meningitidis* serogroup B strain  
RT MC58";  
RL Science 287:1809-1815(2000).  
DR EMBL: AE002420; AAF62314.1; -.  
DR TIGR: NMB0653; -.  
KW Complete Proteome.  
SQ SEQUENCE 422 AA; 45845 MW; B5428F16A296585F CRC64;

```

alignment_scores:      Length: 464
                        Quality: 658.00
                        Ratio: 2.419
                        Gaps: 10
Percent Similarity: 58.621  Percent Identity: 33.405

alignment_block:
US-09-303-518D-463 x Q9K0F4 ..

Align seg 1/1 to: Q9K0F4 from: 1 to: 422

```

[illegible]

```

71 eGlnAspTyrThrHisGlnMetGlyAsnLeuLeuIleGlnGlnAlaAsnI 88
251 TTGAGGAATATCGGTACATGTGCGCTTTCCGATTCACGGCCACAAA 300
88 IeAsnGlyThrIleGlyTyrHisThrArgPheSerGlyHisGly 104
301 TTCCATTGCGCCCTTGACACACATGCGTCACATTCGATTCCTGTGACGAGC 350
105 GlnHisAlaProPheAspAsnHisAlaAlaAspSerAlaSerGlyGly 121
351 CGGATGTCCTGACGATTCACGCTTTACCGCATCCATTGGACGCGAT 400
121 sGlyAsnValAspGlyGlyPheThrValTyrArgLeuAsnTrpLeuGly 138
401 AGCAACACATCCCGCGCGCGATGACGCGGACGCGGCGGCGGCGTAT 450
138 IAsGlnHisHisProAlaAspAlaValTyrArgProGlyGlyAsnTyr 154
451 CCCGCTCCCAAGGCGGAGGATATATACAGTACGACATAAAGGCGT 500
155 ProGlySerProThrGlyAlaArgAspGlyTyrThrHisValAsnGly 171
501 TGCCCAAAATATCGGCTCAACCTGACCGGACGACGACGACGACGAC 550
171 rAlaArgSerIleLeuAsnProThrAspThrArgSerIleArgGlnAla 188
551 GCGTTGCGGACGCTTTCCACATGCGCGGCTATGCTGACGACGAGGAT 600
188 rGlyLeuSerAspAsnTyrSerAsnLeuGlySerAsnPheSerAlaGly 204
601 GGCACGCGATTCAACGCGCGACCGCGATACAGCGCGGCGGCGGACATC 650
205 AspGlyAlaAsnArgGlyMetTrpGlnHisAlaLysLeuAspArgTyr 221
651 GGGCAATGCGCGGAGCGCTTCAACGCGGACGATGATGTCAAAAACA 700
221 pGlyAsnSerMetGlnPheIleAsnGlyAlaAla..... 232
701 TCATCGGCGCGGCGAGAAATGTCGCGCGGCGGCGATGCGCGAGGAT 750
233 .AlaGlyAlaLeuAsnProPheIleSerIleGlyAlaValAlaAsp... 247
751 ATAAAGCAAGGCTCAACATGCTGTCAATGACGCGCTTGCGTCTTTC 800
247 ..... 247
801 CACCGAAACAGATGCGCGCATCAACGATTTGGCAGATATGCGCCAC 850
247 ..... 247
851 TCAAAAGCTATGCGCGAGCAGCATCCGATGCGGAGTTCGCAAAACCC 900
248 .....GlnTrpMetGlnGlnAsnPro 254
901 AATGCGGCAAGGCGATAGAGCGCTGACGATATCTTTATGGCAGCCAT 950
255 AsnAlaIleGlnThrValGlnAlaLeuValAsnValLeuProPheAla... 270
951 CCCCATCAAGGATGAGAGCTGTCGCGGAAATACGCGTTGGGCGGA 1000
270 ..... 270
1001 TCACGCGACATCTGTCAAGCGGTGCGAGATGCGCGCATGCGATTGCG 1050
271 .....LysValLysAsnLeuThrLysAlaAlaLysPro 281
1051 AAAGGAAATCCGCGGTGACGAGCATTTTGGCGATGCGGCGATACGCCAA 1100
282 ...GlyLysAlaIleValSerGlyAspPheSerAspSer..... 293
1101 ATACCGGTCCCTTACCATTCGCCGAATATCCGTTCAAACTTGAGACGC 1150
.....

```

```

294 .....TyrLysHisAsnThrAlaSerArgLeuSerGln 305
1151 GTTACGGCAAGAAACATCACCTCTCAACCGTCCGCGGTCAACGCGC 1200
305 eValAspGlyGlnMetPheGlnThrArgAsnVal.....AspPhe 318
1201 AAAAATGTCAAACTGGCAGACCAACGCCAACCCGAGACAGCGGTACCGTT 1250
319 LysAlaLysSerIleGlyThrLysIleHis..... 328
1251 TGACGGTAAAGGCTTCCGAATTTTGAGAAGCAGCGTAATAT..... 1293
329 .AspGlyAlaGlnGlyLysHisIleSerGlyHisArgAsnTyrIleGln 345
1294 .....GATACGAAGCTCGATATTCACATATATCGGGGCGC 1329
345 LysSerThrLeuAsnGlnAsnIleAsnProGlnGlyLeuLeuAsnGly 361
1330 GGATATCTAAGGCTGATGCTGTTGATGGCAACCGAGA 1371
362 IleHisSerGlyAlaTyrProValIleSerLysGlyAlaArg 375

seq.name: sp_bacteria:Q9X6P3
seq_documentation_block:
ID Q9X6P3 PRELIMINARY; PRT; 509 AA.
AC Q9X6P3:
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ADHESIN MAFB.
GN MAFB.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS11;
RA Eickert-Jeager S., Meyer T.F., Fischer E., Mäler J., Manning P.A.,
RL Rudel T., Schuering-Pflug I., Schulz E., Schwan E.T.;
DR EMBL; AF142582; MAD31039.1;
DR EMBL; AF142582; MAD31039.1;
SQ SEQUENCE 509 AA; 55681 MW; 88B420CFE8269D4 CRC64;

alignment_scores:
Quality: 450.00 Length: 417
Ratio: 1.875 Gaps: 18
Percent Similarity: 57.554 Percent Identity: 32.374

alignment_block:
US-09-303-518D-463 x Q9X6P3 ..
Align seg 1/1 to: Q9X6P3 from: 1 to: 509

31 CTGTCCATATCGGCGATGTCGCTGCGCGATGATGACGACGCGTCAGATT 80
16 IleSerLeuLeuGlnIleProIle.....SerHisAlaAsnGlyLe 29
81 GGCAGGATCCCTTATCGCGGAGTTCTGACCGTCGACGATTTGCAAC 130
29 u.....AspAlaArgLeuArgAspAspMetGlnAlaLysHisTyrGln 44
131 CCGAGCGGAATACCACTATTCGCG..AGCAGGGGGAGCTTGCCNAG 177
44 rGlyGlyLysTyrHisLeuPheGlyAsnAlaArgGlySerVal..... 58
178 CGCAACGCGCATATCGGATTTGGAAACATACAAAGCCATGATGGGCGCA 227
59 LysAsnArgValCysAlaValGlnThrPheAspAlaThrAlaValGlyPr 75
228 CCGTATG.....ATTCAGAGGCGCGCGTTGAAGAAATATCGCGT 268
.....

```





```

319 AACATGCGCTCATTCGATTCGACGAGCGCGTAGTCCCGTTGACGG 368
||||| ||| ||| ||||| :|||: |||
109 AsnHisAspSerLysSerThrSerAspPheSerGlyValAspGlyI 125
|||||:|||||:|||||:|||||:|||||:|||||
369 ATTGACGCTTACCGCATTCATGGGAGATAGCAACACCATCCGCGG 418
|||||:|||||:|||||:|||||:|||||:|||||
125 yPheThrValTyrGlnLeuHisArgThrGlySerGlnLeuHisProGlu 142
419 ACGGTATGAGCGGCGACAGCGCGCGCTATCCCGCTCCCAAGCGCG 468
|||||:|||||:|||||:|||||:|||||:|||||
142 spGlyTyrAspGlyProGlnGlySerAspTyrProProGlyGlyAla 158
469 AGGATATATACAGTACGACATATAAAGCGCTGCCCCAAATATCCGCT 518
|||||:|||||:|||||:|||||:|||||:|||||
159 ArgAspIleTyrSerTyrTyrValLysGlyThrSerThrLysThrLysSe 175
519 CAACCTGACCGCAACCGCGACCGCGCAACAGCGCTTGGCGACCGTTTC 567
|||||:|||||:|||||:|||||:|||||:|||||
175 rAsnIleValProArgAlaPro.....PheSerAspArgTrpL 188
568 .....CACATGCGCGCGCTATGCTGACGCAAGAGATAGCGAGCGATTC 612
|||||:|||||:|||||:|||||:|||||:|||||
188 euLysGluAsnAlaGlyAla.....AlaSerGlyPhe 198
613 .....AACGCGCGCACCGCATACAGCCCGCA 638
199 PheSerArgAlaAspGluAlaGlyLysLeuIleTrpGlnSerAspProAs 215
639 GCTGAGACAGATCGGCGCAATCGCGCGAGACCTTCACGCGACTGCAGATA 688
|||||:|||||:|||||:|||||:|||||:|||||
215 nLysAsnTrpTrpAlaAsnArgMetAspAspIleArgGlyIleValGlnG 232
689 TCGTCAAAAC.....ATCATCGCGCGCGCGAGAGAAATTCGCGCGCA 732
|||||:|||||:|||||:|||||:|||||:|||||
232 LysAlaValAsnProPheLeuMetCylGheGlnGlyAlaGlyIleGlyAla 248
733 .....GGCGATCCGTCAGAGGTATAGCGAAAGCGCTCAAAATTCGCTG 776
|||||:|||||:|||||:|||||:|||||:|||||
249 IleThrAspSerAlaValSerProValThrAspThrAlaAlaGlnGlnTh 265
777 CATCAGACGCG.....TTGGTCTGCTTCACCGGAAACAAAGATGG 817
|||||:|||||:|||||:|||||:|||||:|||||
265 rLeuGlnGlyIleAsnHisLeuGlnAsnLeuSerProGlnAlaGlnLeu 282
818 CGCGCATCAACGATTTGGCAGATATGGCG...CAACTCAAGACATATGCC 864
|||||:|||||:|||||:|||||:|||||:|||||
282 laaIaAlaThrAlaLeuGlnAspSerAlaPheAlaValLysAspGlyIle 298
865 GCAGCAGCCATCCGCGATTTGGCAGATCCCAAAACCGCATCCGCGACAGG 914
|||||:|||||:|||||:|||||:|||||:|||||
299 AsnSerAla...ArgGlnTrpAlaAspAlaHisProAsnIleThrAlaTh 314
915 CATAGAGACCGCTCAGCAATCTTATGGCAGCGATCCCGCATCAAGGG 963
|||||:|||||:|||||:|||||:|||||:|||||
314 rAlaGlnThrAlaLeuAlaValAlaGlnAlaAlaThrThrValTrpGly 331
963 ..... 963
331 LysLysValGlnLeuAsnProThrLysTrpAspTrpValLysAsnThr 347
964 .....ATTGAGCTGTCCGGGAAATACGCTTGGCGCGCATTCAC 1004
|||||:|||||:|||||:|||||:|||||:|||||
348 GLyTyrLysThrProAlaValArg.....Th 356
1005 GGCACATCCTGTCAAGCGGTGCGAGATGGCGGATCGCATTCGCGAAG 1054
|||||:|||||:|||||:|||||:|||||:|||||
356 rMetHisThrLeuAspGlyLumetalGlyLysAsnArgProPro.... 371
1055 GGAATCCCGCGCTCAGCGACATTTTCCGATCGG..... 1089
|||||:|||||:|||||:|||||:|||||:|||||
372 ..LysSerIleThrSerAsnSerLysAlaAspAlaSerThrGlnProSer 387

```

```

1090 .....GCATACGC 1097
388 LeuGlnAlaGlnLeuIleGlyGlnGlnIleSerSerGlyHisAlaTyrAs 404
1098 CAATATC.....CCGTCGCCCTTACCATTCCTCCCAATATCCGTCAA 1138
|||||:|||||:|||||:|||||:|||||:|||||
404 nLysHisValIleArgGlnGlnIleGlnIlePheThrAspLeuAsnIleAsnSerP 421
1139 ACTTGGAGCAGCGCTTACGCGCAAGAAACATCACTCTCAACCGTGGCG 1188
|||||:|||||:|||||:|||||:|||||:|||||
421 rAlaAspPheAlaArgHisIleGlyLysAlaIleValSerHis..... 434
1189 CCGTCAACGCGCAAAATGTC 1209
|||||:|||||:|||||:|||||:|||||:|||||
435 ProThrAsnMetLysGlnLeu 441

seq_name: sp_bacteriap:Q9JXD2

seq_documentation_block:
ID Q9JXD2 PRELIMINARY; PRT; 483 AA.
AC Q9JXD2;
DT 01-OCT-2000 (TRENBLREL. 15, Created)
DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)
DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
DE MARB PROTEIN.
GN NMB2105.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBT_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B.
RX MEDLINE=201755; PubMed=10710307;
RA Tettelein J., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Uitterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlata V., Maignani V., Pliza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappapoli R., Venter J.C.;
RT *Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.
RL Science 287:1809-1815(2000).
DR EMBL; AF002559; AAF62340.1; -.
KW Complete proteome.
SQ SEQUENCE 483 AA; 52789 MW; D6B5AEFA8AEBA02 CRC64;

alignment_scores:
Quality: 390.00 Length: 433
Percent Similarity: 55.658 Percent Identity: 30.716

alignment_block:
US-09-303-518D-463 x Q9JXD2 ..

Align seg 1/1 to: Q9JXD2 from: 1 to: 483

31 CTGTCCATACGTGGCAGTGTGCTGCGCATGCATGCACAGCCCTGAGATT 80
|||||:|||||:|||||:|||||:|||||:|||||
16 IleSerLeuLeuGlnIleProIle.....SerHisAlaAsnGlyLe 29
81 GGCAGAACGATCCCTTATCCGCGCAGGTCTGACCGTCGACATTTGAGAC 130
|||||:|||||:|||||:|||||:|||||:|||||
29 u.....AspAlaArgLeuArgAspAspMetGlnAlaLysHisTyrGlnP 44
131 CGAGCGGGAATATACACCTATTTCGCG...ACGAGGGGAGGCTTGCCNAG 177
|||||:|||||:|||||:|||||:|||||:|||||
44 rGlyGlyLysTyrHisLeuPheGlyAsnAlaArgGlySerValLysLys 60
178 CGCAAGCGCATATTCGATTGGGAACATACAAAGCCATCAG..... 219

```



```
263 TCGGCTACATTGTCGCTTTCCGATCGAGGCGCAAAATTCATTGCGCC 312
||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
291 gATgLIeHsSerPro...PheArgSerArgSerArgSerProLLeArga 307
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
313 TT...CGACAAACATGCTCGATTCGATTCGACGAAGCCGCTAGTCC 359
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
307 rGHIsArgrArgrProthrHIsGluGlyArgArgGlnSerProAlaProSer 323
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
360 CGTTGACGAGATTCAACCTTACCGCATCCATTGGAGCGGATACGAACACC 409
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
324 ArgArGArGArGArSerProSerPro...Pr 332
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
410 ATCCCGCGGAGCGCTATGACGGCGGCGGCTATCCCGCTCC 459
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
332 oAlArGrArGr...ArgSerProSerProProa 343
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
460 AAAGCGCGGAGGATATATACGCTACGACATAAAGCGCTGCCGCAAA 509
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
343 laArGrArGr...ArgArgSerProSer 350
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
510 TATCCGCTCAACCTGACCGACACCG...CAGCACCG 544
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
351 ProProAlaArGrArGHIrArGrSerProThrProProAlaArgGlnArGr 367
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
545 GACAAAGCGCTTGGCGACCGCTTCCAAATCGCGGCTATGTCGACGCA 594
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
367 gSer...ProSerProProAlaArGrArGHIrArGrSerPro 380
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
595 GAGTAGCGGAGGATTCAAACGGCGGCGGATACGCCGCCGAGCTGGA 644
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
380 roProAlaArGrArGr...ArgArgSerProSerProPro 391
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
645 CAGATCGGCGAATCCCGCGAGCGCTTCAAGCGGCTACGATATGCTCA 694
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
392 ...AlaArGrArGrArGrSerProSerProProAlaArGrArGr 406
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
695 AAAACATCATCGCGCGCGGAGGAAT...TGTC 726
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
406 gSerProSerProLeuThrArGrArGrArGrSerProSerProLeuThr 423
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
727 GCGCGACGCGATGCGGCTGACGGGTATTAAGGAGGCTCAAACTGCTGT 776
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
423 rGrArGrArGrArGrSerProLeuAlaArGrArGrArGrSerArGr 439
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
777 CATGCAAGCGCTTGGCTGCTTCCACCGAAACAGAT...GGCGC 820
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
440 ProGlyArGrSerProSerProValAlaArGrLeuArGrAspProThrGly 456
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
821 G...CATCAAGATTTGGCAGATATGGCGCA 849
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
456 ArGrLeuProSerProSerIleGlnArGrLeuProSerProProVal 473
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
850 CTCAAAAGCTATGCGCGACAGCCATCCGCAATGGGAGTCGCAAAACC 899
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
473 laGlnArGrLeuProSerProProArGrArGrArGrLeuProSerPro 489
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
900 CAATGGCGGACAAAGCATAGAAC... 923
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
490 ProProAlaGlnArGrLeuProSerProProProArGrArGrArGrLeu 506
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
924 .CGTCGCAATATCTTTATGCGAGCAT...C 951
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
506 oSerProMetArGrIleGlySerHIsAlaAlaSerHIsLeuGlnSer 523
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
952 CCCATAAAGGATTTGGAGCTGTCGGGGAATAACGGCTT... 992
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
523 roSerProSerSerLeuSerProProGlyArGrLeuValLeuProSer 539
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
992 ... 992
540 ProProValArGrArGrArGrSerLeuThrProAspGlnGlyArGrVal 556
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
993 ...GGCGGCGATCAGCGGACATCTGTCAAGCGGTCCGAG 1029
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
556 rLeuSerGlnGlyArGrHIs...ThrSer...ProSerHIsLeuGln 571
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1030 ATGAGCGCGGATCGCA...TTCGCAAAAGGCAAAATCCGCGCTCGCA 1073
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
572 AspGlySerMetSerProValArGrGlyArGrGlyLysSerProSer 588
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1074 CAATTTTGGCGATCG...GCATACGCGCAAAATCCGCTCCCTT 1114
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
588 rArGrHIsGlnLysAlaArGrSerProValArGrArGrArGrSerProThrPro 605
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1115 ACCATTCGCGAAATATCCGTTCAAACTTGGACAGCGT...TAC 1155
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
605 AlaSerArGrArGrSerArGrSerSerArGrSerArGrSerProAspArGr 621
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1156 GCGCAAAAGAAACATCACTCTCTCAACCGCGCGGCTCAAGGCGCAAAA 1205
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
622 ArgArGrArGrArGrSerProSerSerSerArGrSerProSerArGrSer 638
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1206 TGTCAAACTG...G 1216
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
638 rProProValLeuHIsArGrSerProSerProArGrArGrArGrHIsGln 655
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1217 CAGACCAAGCGCACCG... 1233
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
655 rGlnArGrArGrSerProGlyArGrLeuSerGlnGlnAspArGrValGln 671
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1234 ...AAGACAGCGCTACCGCTTTCAGCGTAAAGGCTT 1265
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
672 AsnSerLysLeuLeuLysArGrThrSerVal... 681
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1266 TCCGATTTTGAAGAGCAGTGAATATGATATCGAAGCTCGATTTCAG 1315
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
682 .ProAspThrAspLysArGrLysGlnLeuProGlnLysLeu... 694
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1316 AATTATCGGCGGCGGCTATACCTTAAGCGTACCGCTTTGATGACGAAA 1365
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
695 ..LeuGlnValGlyArGrValGlnHIsTyrLys... 704
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1366 CCGACATGGAGGAGTTGATAGCAAGCTTAATAATTGACACTCGTGACA 1415
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
705 ...GlnGlnLysArGrLysSerAspLysLeuSerGlnLysArgSer 718
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1416 GGTGGAAGAAAT...GTTACGCAAA 1438
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
718 rValHIsArGrHIsHIsGlySerGlnMetSerProValGlnAsnSerGln 735
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1439 CGAAGAAGAGAGTCAGAGTAGTCAGTTAAAGCCATGCGCAACAGAGA 1488
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
735 LysArGrSerArGrProValSerSerLysValLysAspSerGlnGlnValGln 751
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1489 TGGCAAAATTAACAGGCTTAGATTATTCATTTTATAGGTGATAT 1558
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
752 LysGlnAspAsnSerAspLeuAspAlaAsnLeu...SerCy 764
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1539 CAATAGAAAGGACAGTAAACAGAGGCGCATAGTCA...A 1576
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
764 sAspSerLysAspThrIleArg...HIsGlnIleLysAspLysAsnA 1779
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1577 CCCGTTGATGATACGGGTATACAAACAAACCTGGGACCTGATAAACAT 1626
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
779 rGrArGlyAsnLysArGrSerSerArGrGlnValSerSerAspAspAsn 795
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1627 GGGGTTTATCAAGGACAGTGGAAATTAAAGCTGATGGAAGCTGGGA 1676
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
796 GlySerSerAspSerAspValAspAspArGrLys...G1 807
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1677 GGTGAAGAAAGAAAGAGTGGAAAGTGAATGACCAAG 1713
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
807 uAlaLysArGrArGrGlyGlnGlnLysLysThrArGr 819
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

seq\_name: sp\_plant:092W08

seq\_documentation\_block:  
 ID Q92W08 PRELIMINARY; PRT: 891 AA.  
 AC Q92W08;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE PUTATIVE PROLINE-RICH PROTEIN.  
 GN AT2G29210.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
 RA Buehl C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., Vanden S.E., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhaver G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,  
 RA Salberg S.L., Fraser C.M., Venter J.C.;  
 RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana";  
 RL Nature 402:761-768 (1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC004561: AAC95214.1; -  
 DR InterPro: IPR002483; PWT.  
 DR InterPro: IPR002965; P\_Lich\_extensn.  
 DR Pfam: PF01480; PWT; 1.  
 DR PRINTS: PR01217; PRICEXTENSN.  
 DR SMART: SM00311; PWT; 1.  
 SO SEQUENCE 891 AA; 102261 MW; C0CGB6DE055E528B CRC64;

alignment\_scores:  
 Quality: 163.50 Length: 599  
 Ratio: 0.654 Gaps: 30  
 Percent Similarity: 41.736 Percent Identity: 23.706

alignment\_block:  
 US-09-303-518D-463 x Q92W08 ..  
 Align seg 1/1 to: Q92W08 from: 1 to: 891

```

255 AGGAATATGGGCTACATTCGCGCTTTCCGATCAAGGCGGACAAATTCG 304
||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
286 ArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 301
305 ATTGGCCCTT...CGACACCATGCTCATTCGATTCGATTCGACGAAGCC 351
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 cileArgArgArgArgArgArgArgArgArgArgArgArgArgArg 318
352 GGTAGTCCCTTGACGATTCAGCTTTACCGATCCATTCGATTCGACGATA 401
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
318 lArProSerArgArgArgArgArgArgArgArgArgArgArgArg 328
402 CGAACACCATCCCGCGGCGGCTATGACGGGCGGCGGCGGCTATC 451
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
329 .....ProAlaArgArgArg.....ArgSerProse 337
452 CCGCTCCCAAGGCGGAGGATATATACAGCTACGACATTAAGGCGTT 501
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
337 rProProAlaArgArg.....ArgArgS 345
502 GCCCAAAATATCGGCTCAACCTGACGCAACGCG..... 536
||||| ||| ||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

345 erProSerProProAlaArgArgHisArgSerProThrProProAlaArg 361
537 CAGACCGGACGACAGCGCTTCCGACCGCTTCCAAATGCCGCGCTATGC 586
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
362 GlArgArgArgSer.....ProSerProProAlaArgArgHisArg 374
587 TGACGCAAGAGTAGCGGACGATTAACAGCGCCACCGCATACAGCCC 636
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
374 gSerProProProAlaArg.....ArgArgSerProSerProPro. 388
637 GAGCTGACAGATGCGGCAATGCCGACGCTTCAACGCGCACTGACAGA 686
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
389 .....AlaArgArgArgArgSerProSerProProAlaArg 400
687 TATGTCAAAACATCATCGCGCGGCGAGAAAT..... 722
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
401 ArgArgArgSerProSerProLeuTyrArgArgAsnArgSerProSerPr 417
723 .....TGTCGCGGACGCGATGCCGTGACGGGTATAGCGAAGCTCAAC 768
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
417 OleuTyrArgArgAsnArgSerArgSerProLeuAlaLysArgGlyArgS 434
769 ATTGCTGTCATGACGCGCTTGGCTGCTGCTTCCACGCAACAAAGAT... 815
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
434 erAspSerProGlyArgSerProSerProValAlaArgLeuArgAspPro 450
816 ..GCGCGG.....CATCACGATTTGGCAGATA 841
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
451 ThrGlyAlaArgLeuProSerProSerIleGluGlnArgLeuProSerPr 467
842 TGCGGCACTCAAGACTATGCCGACGACGACGCGGATTTGGGCGAGTC 891
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
467 oProValAlaGlnArgLeuProSerProProAlaArgArgGlyArgLeuP 484
892 CAAACCCCAATGCCGCGACAGCATAGAAC..... 923
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
484 roSerProProProAlaGlnArgLeuProSerProProAlaArgArgAla 500
924 .....GTCAGCAATATCTTTATGCGAGCAT..... 950
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
501 GlyLeuProSerProMetArgIleGlyGlySerHisAlaAlaAsnHisLe 517
951 .....CCCATCAAGGAGATGAGCTGTCGCGGCAAAATACGCGCT. 992
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
517 uGlnSerProSerProSerSerLeuSerProProGlyArgLysLysValL 534
992 ..... 992
534 euProSerProProValArgArgArgSerLeuThrProAspGlu 550
993 .....GGGCGCATCGACGCGCATCTCGTCAAGC 1021
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
551 ArgValSerLeuSerGlnGlyArgHis.ThSer...ProSerHisI 566
1022 GGTCCAGATGGCGCGCATGCCA.....TGCCGAAAGGAAATCCGCC 1065
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
566 lArgGlnAspLysSerMetSerProValArgGlyArgGlyLysSerSer 582
1066 GTCAGCGACAAATTTGCCGATCG.....GCATCGCAAAATATACC 1106
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
583 ProSerSerArgHisGlnLysAlaArgSerProValArgArgArgSerPr 599
1107 GTCCCTTACATCCGCAATATCCGTTCAAACTTGAGAGCAGCT.... 1152
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
599 oHnProValAsnArgArgSerArgArgSerSerSerAlaSerArgSerP 616
1153 .....TACGCAAAAGAAACATCACTCTCAACCGTCCGCGCAAC 1197
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
616 roAspArgArgArgArgSerProSerSerSerArgSerProSerArg 632
1198 GCGCAAAATGTCAAACTG..... 1215
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
633 SerArgSerProProValLeuHisArgSerProSerProArgGlyArg 649

```

```

1216 .....GCAGACCAAGCCACCGC..... 1233
649 SHISLmrglunArgSerProGlyArgLeuSerGluGlnAspA 666
1234 .....AAGACGAGCGTACCGTTGAGCGT 1237
666 rGValGlnAsnSerLysLeuLysArgThrSerVal..... 678
1258 AAGGGTTTCGGAATTTGAGAACAGCTGAATATATATACGAAGTCGA 1307
679 .....ProspThrAspLysArgLysGlnLeuProGluLysLeu.. 691
1308 TATTCAGAAATTATCGGGGCGCGGTATACCTAAGCTAAGCTAGCTGTGTTG 1357
692 .....LeuGluValGlyArgValGlnHisTyrLys..... 701
1358 ATGGGAACCGACAGATGGAGGTGATAGAACTTAATAATGACAACT 1407
702 .....GluGlnGluArgLysSerAspLysLeuSerGlu 712
1408 CGTAGCAGGTGAGAAAAAT.....GT 1430
713 LysArgSerValHisArgHisGlySerGlnMetSerProValGluAs 729
1431 TCAGGAACGAGAAAGAGAGCTAGCTAGCTTAAGCCCATGCGC 1480
729 nSerGluGlyArgSerArgProValSerSerLysValLysAspSerGlu 746
1481 AACGAGATGGGAAATTAAGAGGCTTATATCTTTATTAAGGT 1530
746 InValGluLysGluAspAsnSerAspLeuAspAlaAsnLeu..... 759
1531 GGTGATATCAATAAGAAAGCAGCTAACAGAGCGCATAGTCTA..... 1575
760 .....SerCysAspSerLysAspThrIleArg.....HisGlnIleLysAs 773
1576 .....ACCCGTGTCATGTACGGGTGATACAAACCTCGGACCTG 1618
773 pLysAsnArgArgLysAsnLysArgSerSerArgGluValSerSerA 790
1619 ATAAACATGGGGTTTATCAAGCAGCAGCTGAATTAAGAAAGCTGATGCA 1668
790 spAspAsnGlySerSerAspSerAspValAspAspArgLys..... 803
1669 AGTTGGAGGTGAAGAAAGGTTGGAAAGTATGATGACCAAG 1713
804 .....GluAlaLysArgArgArgLysGluGluLysLysThrArg 816
seq_name: sp_invertebrate:Q9N4S7
seq_documentation_block:
ID Q9N4S7 PRELIMINARY; PRT; 1079 AA.
AC Q9N4S7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DR 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
GN Y51B11A.1 PROTEIN.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;
OC Rhabditidae; Peioderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RT None;
RL "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-BRISTOL N2;
RA Cotton M.;
RT "The sequence of C. elegans cosmid Y51B11A.";
RN Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
DR EMBL: AC006797; AAF60743.1;
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS: PR01217; PRICHEXTENS.
SQ SEQUENCE 1079 AA; 110532 MW; 8BDE3824CF80CA1 CRC64;

```

```

alignment_scores:
  Quality: 162.50      Length: 421
  Ratio: 0.781        Gaps: 20
  Percent Similarity: 49.406   Percent Identity: 24.703

```

## alignment\_block:

US-09-303-518d-463 x Q9N4S7 ..

Align seg 1/1 to: Q9N4S7 from: 1 to: 1079

```

71 CCTCAGATTGGCAACATCCCTTTATTCGCGCAGCTTCTGCACGCTGAG 120
   |||...:|||||  |||  |||...:  |||...
81 ProGluThrThrSerThrGluProSerSerSerThrThrProValGlu 97
121 CATTCGAACCGCGGGAATATACCACTATTCGCGCAGCAGCGGGGAGCT 170
   |||...:|||||  |||  |||...:  |||
97 nThrThrThrThrThrAlaProGluThrThrSerThrGluAlaProSer. 113
171 TCCCAAGCGCAAGCGCATATCGGATTGGAAACATACAAAGCCA..... 215
   |||...:|||||  |||...:  |||
114 ..SerSerThrThrProValGlnThrThrThrThrThrAlaProGluTh 129
216 ...TCAGTGGCGCACCTGATGATTCACAGAGCGCGCTGAAAGAAAT 261
   |||...:|||||  |||...:  |||
129 rThrSerThrGluProSerSerSerSerThrSer..ProValGln..... 143
262 ATCGGCTCATGTCGCTTTCGATCAGCGGCAATATCCATTCGCGC 311
   |||...:|||||  |||...:  |||
144 .ThrThrThrThrThrAlaProGluThrThrSerThrGluAlaProSer 160
312 CTTCGACACCATGCTTACATTCGATTCGACGAAGCGGTAGTCCG 361
   |||...:|||||  |||  |||...:  |||
160 erSerThrThrProValGlnThrThrThrThrThrThrAlaProGluThr 176
362 TTGACGAT.....TCAGCTTACCGCATCCATTCGGAGCGATAGCA 405
   |||...:|||||  |||...:  |||
177 SerThrGluProSerSerSerSerThrSerProValGlnThrThrThr 193
406 CACCATCCCGCGCAGCGCTATAGCGGCGCACAGCGCGCGCTATCCCGC 455
   |||  |||  |||...:  |||
193 rThrAlaProGluThrThrSerThr.....GluProPro 205
456 TCCCAAGCGCGCGGATATATACACTACGACATTAAGCGCTGCC 505
   |||...:|||||  |||...:  |||
205 erSerSerThrThrProValGlnThrThrThrThrThrAlaPro 219
506 AAAATA.....TCCGCTTACCGCATCCGACG..... 529
   |||...:|||||  |||
220 GluThrThrSerThrGluProProSerSerSerThrThrProValGln 236
530 ....ACAACGCGACACCGGAC.....AACGGCTTCGCGACGCTT 566
   |||...:|||||  |||...:  |||
236 rThrThrThrThrAlaProGluThrThrThrThrThrGluSerProSer 253
567 CCACAAATCGCGCGCTATGCTGACGCAAGAGTAGGCGCATTAAC 616
   |||  |||...:  |||
253 erThrThrProVal.....Gln 258

```









```

735 hrAlaThrProSerSerProGlyArgAlaThrLeu...ProVal:750
751 TGGGGCAACTCAAGACTATGCCGAC...CAGCC 873
751 TTTTLeSerThrThrThrThrProThrThrArgGlySerThrValThrPro 767
874 ATCCGCGATT...GGGAGTCCAAACCCGCAATGCCGACAGGCAATGAGA 920
767 oSerSerThrProGlyThrThrThrThrProThrValLeuThrThrThr 784
921 AGCCGTCAGCAATATCTTTATGACAGCCATCCCATCAAGGATTTGAG 970
784 hrThrThrValAlaThrGlySerMetAlaThrProSerSerSerThrGln 800
971 CTGTCCGGGGAATAACGGCTTGGGGGCAATCAGGCAATCTCTGCAAG 1020
801 ThrSer... 802
1021 CGTCCGAGATGGGCGCATGCGATTGCCGAAGGAATCCGCGTCAG 1070
803 .....GlyThrProProSerL 808
1071 CGACATTTTGGGATGCCGATACG...CCA 1099
808 eutThrThrAlaThrThrThrThrThrThrThrThrThrThrThrPro 824
1100 AATACCGCTCCCTTACATCCGCAATATCCGTCATCAACTGGAGCAG 1149
825 SerSerThrProGlyThrThrProThrProThrValLeuThrThrAla 841
1150 CGTTACGGCAAGAAACATCA... 1171
841 aThrThrProAlaAlaThrSerSerThrValThrProSerSerAlaLeuG 858
1172 .....CCTCTCAACGCTGCCGCGCTCAACGCAAAATG 1207
858 lYThrThrThrThrProThrProThrProThrProThrThrAlaThr 871
1208 TCAAACTGGGAGACCAAGCCGACAGAGAGAGAGGCTACCGTT 1250
872 ThrHis.GlyArgSerLeuSerProSerSerProThrThrVal 885

seq_name: sp_bacteria:Q53478

seq_documentation_block:
ID Q53478 PRELIMINARY; PRT; 456 AA.
AC Q53478;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ORF2 PROTEIN.
GN ORF2.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95020543; PubMed=7934834;
RA Facius D., Meyer T.F.;
RT "A novel determinant (coma) essential for natural transformation competence in Neisseria gonorrhoeae and the effect of a coma defect on pilin variation."
RL MOL. Microbiol. 10:699-712(1993).
DR EMBL; S75490; AAB32262.1;
SO SEQUENCE 456 AA; 54380 MW; 30AC7B5CDFB912D5 CRC64;

```

```

alignment_scores:
  Quality: 157.50      Length: 395
  Ratio: 0.890        Gaps: 22
  Percent Similarity: 44.810  Percent Identity: 25.316

```

```

alignment_block:
US-09-303-518d-463 x Q53478
Align seg 1/1 to: Q53478 from: 1 to: 456

```

```

477 ATACAGCTACGACATTAAGAGCGT...TGCCAAATATTCGGCTCAAC 523
121 lIeGlnThrArgHisAspThrArgLeuGlyArgArgArgArgArgGln 137
524 TGACGACACACCGGACGACCGGACCGGCTGGCGGCTTCCACAT 573
137 eAluLysGln...MetAlaValLysArgPro... 146
574 GCGGCGCTATGCTGACGCAAGAGTACGAGCATTCACAAAGCGCCAC 623
147 .....HisLysHisArgGlnAlaLeuProArgLle 156
624 CCGATACAGCCCGAGCTGGACAGATCGGCAATGCCCGAAGCTT... 671
157 ProAsnArgAsnArgSerPheValPheArgTyrCysPheArgThrValG 173
672 .....CAAGGAC...T 681
173 TyrLysProGlyGlnGlnGlnHisArgProArgThrGlnThrGlnAla 190
682 GCAGTATATGTCAAACATCATCGCGGCGAGAGAAATGTCGGCGC 731
190 lArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 206
732 AGCGATCGCGTGCAGGTTAAGGAGGCTCAACATTCG...TG 775
207 GlnArgCysGlyLysGlnGlyLysLeuArg...ArgHisLleGlyLysPr 222
776 TCATGACAGGCTGGGCTGCTTCCAAAGAAACAGATGCGCGCATC 825
222 OHISGlnLysProGlnSerValPheArgGlnLysCysProArgLeuL 239
826 AACGATTTGGCAGATATGCGCAACTCAAGCATATGCGGACAGCCAT 875
239 yAspArgLysArgHisAspArgThrGlnGlnGlnGlnGlnGlnGln 255
876 CCGGATGGGCGACGTCCAAACCCCAATGCCGACAG... 914
256 ProArg...LysProArgTyrArgHisArgLleAspGlnArg 268
914 ..... 914
268 gAlaAspGlnArgGlnArgGlyLysGlnLleThrGlnGlnAsnGlnArg 285
915 ...CATGAGAGCGGTCAGCAATATCTTATGCGACCCATCCCAT... 956
285 lHisArgSerArgProLeuProAlaHisLysGlyLeuProPheProLeu 301
957 ...CAAGGATTTGGAGCTGCTCCGGGGAATAAGCGCTTGGCGGATCA 1003
302 ValGlnThrProCysThrArgProTyrGln...GlyArgHisGln 315
1004 CGGACATCTCTGTACACGGTCCGACAGTGGCGCGCATGCG...ATTGCCG 1050
315 n...ThrGlnArgGlnProSerThrAspThrGlnAsnArgGlnArgValG 331
1051 AAAGGAATCCGCGCTGACGACCAATTTTGGGATCGGCGGCAATACGCCAA 1100
331 lGlnGlnGlnHisArgArgGlnArgLeuProArgArgArgProArgArgGln 347
1101 ...ATACCGCTCCCTTACCATTCGCAATATCCGTTCAAACTTGAGAC 1147
348 SerSerProAlaProArgPro... 354
1148 AGCGTTACGGCAAGAAACATATCATCTCTCAACGCTGCCGCGGTCAAC 1197
355 .....ArgGlnArgArgGlnHis.GlnAsnThr... 363

```



```

644 GlySerArgSerArgThrProAlaArgGlyArgSerArgSer 660
776 TCATCCAGCGCTGGCTGCTCTCCACCGA.....AACACAGATG 816
660 gThProArgArgGlyArgSerArgSerArgSerLeuValArgGly 677
817 GCGGCGCAT.....CAACGATTTGGCAGATATGGCGCA..... 848
677 rGSerHisSerArgThrProGlnArgArgGlyArgSerGlySerSer 693
849 .....ACTCAAGACTATGCGCAGC...AG 871
694 GlnArgLysAnlySerArgThrSerGlnArgArgSerArgSerL 710
872 CCATCCGCA.....T 882
710 rSerProGlnMetLysLysSerArgLleSerSerArgArgSer 727
883 TGGGCACTCCAAAACCCCAATGCGGCACAGCATAGAACCGTCAGCA 932
727 eSerSerProArgSerLysAlaLysSerArg...LeuSerLeuArg 742
933 TATCTTATGGCAGCCAT.....CCCATCAAGAGGATTTGAGCTGTC 976
743 SerLeuSerGlySerSerProCysProLysGlnLysSerGlnThr 759
977 GGGGAAATACGGCTTGGCGGCATCAC.....GGCAGAT 1011
759 oArgArgSerArgSerLysSerSerGlnProLysAlaLysSerArg 775
1012 CCTGTCAAGGCTGCGCAGATGGCGCGCATGCAATTGCCGAAGGAATC 1061
776 ProProArgArgSerArgSerSerSerSerProProLysGlnLys 792
1062 GCGCGTACGACAAATTTGGCGATGGCGATACGCCAATACCCGTC 1111
792 rLysThrProSerArgLysSerHisSerSerSerProHisPro... 807
1112 CTATACATTTCCGAAATATCCGTTCAACTGGAGCAGCTTACGSCAA 1161
808 .....LysValLysSerGlyThrProProArgGlnGly... 818
1162 GAAACATCATCCTCTCAACCGTGGCGCGCTCAACAGCGCAAAATGCA 1211
819 .....SerLleThrSerProGlnAlaAsnGlnLysValInh 831
1212 ACTGAGACACCAAGCCCGGAGACAGCGCTTACCGTTGACGGTAAG 1261
831 r.....ProGlnArgSerCysPheGlnSerSerP 842
1262 GGTTCGCAATTTTGAAGACAGCTGAATATGATACGAGCTCGATTT 1311
842 roAspProGlnLeuLysSerArgThrPro.....SerArg 853
1312 CAAGATTTTTCGGGGCGGCTATACCTAAGCTTACCTGTTGATGTC 1361
854 HisSerCysSerLysSerProProArgValLysSer.....SerTh 868
1362 GAAACCGAGATGGAGTTGATAGAAAGCTTATAAATGACACACGTCG 1411
868 rProProArgLysSerProSerArgSer.....SerSerProG 881
1412 AGCAGTGGAGAAAATGTTCAAGAAAGAGAGAGAGAGTACAGT 1458
881 InProLysValLysAlaLleIleSerProArgGlnArgSerHis 896
seq_name: sp_human:09U040
seq_documentation_block:
ID 09U040 PRELIMINARY; PRT; 1262 AA.
AC 09U040;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)

```

```

DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
DE RNA BINDING PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtaki S., Umeki K., Sawada Y.;
RT "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016087; BA83713.1;
DR InterPro; IPR002965; P_fich_extensn.
DR PRINTS; PR01217; PRICHEXTENSN.
FT NON_TER 1
FT NON_TER 1262
FT SEQUENCE 1262 AA; 138003 MW; 75E76BD73FA14295 CRC64;

```

```

alignment_scores:
  Quality: 157.00      Length: 617
  Ratio: 0.584        Gaps: 30
  Percent Similarity: 43.598      Percent Identity: 22.690

```

alignment\_block:

US-09-303-518D-463 x 09U040 ..

Align seg 1/1 to: 09U040 from: 1 to: 1262

```

11 CCCGCAAAATATCCCTTATTTGTCATACAGCGCAGTGGCTGCCGATG 60
154 ProAlaProThrProLeuLeu.....AlaGlnArgHis 164
61 CATGCACAGCCCTCAGATTGGCAACGATCCCTTATCCGCGCAGG...T 107
164 sGlyLysSerProGlnProLeuAlaThrThrProLeuSerGlnProV 181
108 TCTGCAGCGTACGATTTTCGAACCGGAGGGGAAATACAC... 148
181 alaAsnProSerSerGlnAlaSerProThrArgAspArgSerProLys 197
149 .....TATTCGCGCAGCAGGGGAGCTTGGCCAG 177
198 SerProGlnLysLeuProGlnSerSerSerSerGlnSerProProSe 214
178 CGCAAGCGCCATATCGGATTTGGAAACATACA..... 209
214 r.ProGlnProThrLysValSerArgHisAlaSerSerProGlnSer 230
210 ..AAGCATCATGTTGGGCA..... 227
221 ProLysProAlaProAlaProGlnSerHisArgGlnLysSerSerP 247
228 .....CTGATGATTCACACAGCGCGCGCTTGAAGAAATATCGCTACAT 272
247 cThrSerLysAsnArgSerHisGlyArgAlaLysArgAspLysSerHis 264
273 ..... 275
264 eHisThrProSerArgArgMetGlyArgSerArgSerProAlaThrAla 280
276 CCGCTTTCCGATCAGCGGCGCAAAATTCATTC..... 308
281 LysArgGlyArgSerArgSerArgThrProThrLysArgGlnHisSer 297
309 .....GCCCTTGACAAACCATGCTTCATTCGATTCGATTCTGAC 345
297 gSerArgSerProGlnThrProArgArgSerArgSerAlaGlnArgTrpGly 314
346 GAACCGCGTAGTCCCGTTGACGAGATTACGCTTTACCGCATTCATTGGA 395
314 rGSerArg.....SerProGlnArgArgGly 322

```

```

396 CGGATACGACACATCCCGC..... 416
323 ArgSerArgSerProGlnArgProGlyThrSerArgSerArgAsnThrG1 339
417 .....CGACGC 423
339 nArgArgGlyArgSerArgSerAlaArgArgGlyArgSerHisSerArgS 356
424 TATGAGGCGCACAGGCGCGGTATCCGCTCCCAAGCGCGAGGA 473
356 erProAlaThrArgGlyArgSerArgSerArgThrProAlaArgGly 372
474 TATATACAGTACACATAAAGCGCTTGCCTCCCAATATCCGCCACACC 523
373 .....ArgSerArgSerArgThrProAlaArgArgSerArg 385
524 TGACCGACAACCGCAGCACCGCAACGGCTTGCCGA...CCGTTTCCAC 570
385 gSerArgThrProThrArgArgArgSerArgSerArgThrProAlaArg 402
571 AATGCCGCGCTATGCTGACGCAAGAGTAGCGCAGATTCAACGCC 620
402 rgGlyArgSerArgThrProAlaArgArgSerArgThrArg 418
621 CACCCGATACGCCCGAGCTGACAGATCGGCATGCGCGGAGCGCT 670
419 SerProValArgArgSerArgSerArgSerProAlaArgArgSerG1 435
671 TCAA.....CGGCACGTGACAGATATCGTCAAAA. 698
435 yArgSerArgSerArgThrProAlaArgArgGlyArgSerArgSerArgT 452
699 ..CATCATCGCGCGCGAGCAAAATTGTCG..... 728
452 hrProAlaArgArgGlyArgSerArgSerArgThrProAlaArgSer 468
729 ...CGCAGGAGTGCGCGTACAGGTTAAAGCAGCTCAACATGCTG 775
469 G1yArgSerArgSerArgThrProAlaArgGlyArgSerArgSer 485
776 TCATGACGCGCTTGCTGCTTCCACGCA.....AACAAGANG 816
485 gThrProArgArgGlyArgSerArgSerArgSerLeuAlaArgGlyA 502
817 CGCGGCAT.....CAAGATTGGCAGATATGCGCA..... 848
502 rgSerHisSerArgThrProGlnArgGlyArgSerGlySerSer 518
849 .....ACTCAAGACTATGCGGACG...AG 871
519 G1uArgLyAsnLySerArgThrSerGlnArgArgSerArgSerAnse 535
872 CCATCCGCA.....T 882
535 rSerProGlnLeuLySerArgGlySerSerArgArgSerArgSerL 552
883 TGGCGAGTCCAAACCCCAATGCGCAAGGATAGAACCCGTCAGCA 932
552 euserSerProArgSerLyAlaLySerArg...LeuSerLeuArgArg 567
933 TATCTTATGCGACCAT.....CCCCATCAAGAGGATTGGAGCTTCC 976
568 serLeuSerGlySerSerProCysProLySlnSerGlnProPro 584
977 GCGGAATATACGGCTGCGCGGCGCATC.....GCGCAT 1011
584 oArgArgSerArgSerGlySerSerGlnProLyAlaLySerArgThr 600
1012 CCTGTCAGCGCTGCGAGATGGCGCGATGCGCATTTGCCAAGGAATC 1061
601 ProProAlaArgSerArgSerSerSerProProLySlnLyse 617
1062 CGCGCTACGACAAATTTGGCGATGGCGATACGCCAATACCGCTGCC 1111

```

```

1 ..... 632
617 rLySThrProSerArgGlnSerHisSerSerSerProHisPro.... 632
1112 CTATACCATTCGCCAATATCCGTTCAACTTGAGACAGCGTTCGGCAA 1161
633 .....LysValLySerGlyThrProProArgGlnGly... 643
1162 GAANAATACACTCTCCACCGTCCGCGCTCAACCGCAAAATGTCAA 1211
644 .....SerLeuThrSerProGlnAlaAsnGlnInsValTh 656
1212 ACTGCGACACACACGCCGCCGGAAGACAGCGTACCGTTGACGTTAA 1261
656 r.....ProGlnArgArgSerCysPheGlnSerSer 667
1262 GGTTCGCAATTTGACAGACGCGTGAATATGATACGAAGCTGATAT 1311
667 roaSPProGlnLeuLySerArgThrPro.....SerArg 678
1312 CAAGATTTATCGGGGCGGTATACCTAAGGCTAAGCGCTGTGTGATGC 1361
679 HisSerCysSerGlySerSerProProArgValLySer.....SerTh 693
1362 GAACCGAGATGGAGGTTGATAGGAAGCTTAATAAATTGACAACTGCTG 1411
693 rProProArgGlnSerProSerArgSer.....SerSerProG 706
1412 AGCAGGTGAGAAAATGTTCAAGAACAGAGAGAGTCAAGT 1458
706 InProLyValLyAlaLeuSerProAlaArgGlnArgSerHisSer 721

```

seq\_name: sp\_human:Q9UHA8

seq\_documentation block:

```

ID Q9UHA8 PRELIMINARY; PRT: 2296 AA.
AC Q9UHA8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SPLICING COACTIVATOR SUBUNIT SRM300.
GN SRM300.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20132238; PubMed=10668804;
RA Blencowe B.J., Bauren G., Eldridge A.G., Issner R., Nickerson J.A.,
RA Rosonina E., Sharp P.A.;
RT "The SRM160/300 splicing coactivator subunits.";
RL RNA 6:111-120(2000).
DR EMBL: AF201422; AAF21439.1;
DR InterPro: IPR002965; P_Rich_extensn.
DR PRINTS: PRO1217; PRICHEXTENS.
SQ SEQUENCE 2296 AA; 251964 MW; 17C0BD4EA10A9CF9 CRC64;

```

alignment\_scores:

```

Quality: 157.00 Length: 617
Ratio: 0.584 Gaps: 30
Percent Similarity: 43.598 Percent Identity: 22.690

```

alignment\_block:

US-09-303-518D-463 x Q9UHA8

Align seg 1/1 to: Q9UHA8 from: 1 to: 2296

```

11 CCGCAAAATATCCCTTATCTGTCCATCTGCGAGTGTGCTGCCGATG 60
||||| ||||| ||||| ||||| |||||
364 ProAlaProThrProLeu.....AlaGlnArgHis 374
61 CATGACAGCGCTGACGATTTGGCAAGATCCCTTATCCGCGAGG...T 107

```

```
374 sGlySerProGlnProLeuAlaThrThrProLeuSerGlnGluProV 391
    ::::|
108 TCTCCAGCGTCAGCATTTGCAACCCGAGGAAATATCCACC..... 148
    ::|
391 alaInProSerGlnAlaSerProThrArgAspArgSerProProlys 407
149 .....TATTCGGCAGCAGGGGGAGCTTGCACAG 177
408 SerProGlnLysLeuProGlnSerSerSerGlnSerProProse 424
178 CGCAAGCGCCATATCGAGTTGGAAACATACA..... 209
    ::|
424 r.ProGlnProThrLysValSerArgHisAlaSerSerSerProGlnSer 440
210 ...AAGCCATGAGTTGGGCA..... 227
441 ProLysProAlaProAlaProGlnSerHisArgLysLeuSerSerPr 457
228 ...CCTGATGATTCCACAGCGCGCGCTTGAAGAAATATCGGCTACAT 272
457 OThrSerLysAsnArgSerHisGlyArgAlaLysArgAspLysSerHis 474
273 ..... 275
474 eRhIsthrProSerArgArgMetGlyArgSerArgSerProAlaThrAla 490
276 CCGCTTTCCGATCAGCGGACCAAAATTCATTC..... 308
491 LysArgGlyArgSerArgSerArgThrProThrLysArgLysHisSerAr 507
309 .....GCCCTTCGACCAACCATGCTCCATTCGATTCGATTCGAC 345
507 gSerArgSerProGlnThrParArgSerArgSerArgSerArgThrProLys 524
346 GAAGCGGTAAGTCCCGTTGACGAGATTCAGCCTTATCCGCATCCATTCGGA 395
524 rSerArg.....SerProGlnArgArgGly 532
396 CGGATACGACACCATCCCGC..... 416
533 ArgSerArgSerProGlnArgProGlyTyrSerArgSerArgSntHrGly 549
417 .....CGACGCG 423
549 naArgArgLysArgSerArgSerAlaArgArgLysArgSerHisSerArgS 566
424 TATGACGGCGCACAGGGCGGTATCCGCTCCCAAGGCCGAGGGA 473
566 ePrProAlaThrArgGlyArgSerArgSerArgThrProAlaArgArgGly 582
474 TATATACAGCTACGACATAAAGCGCTTGCCCAAAATATCCGCTCAAC 523
583 .....ArgSerArgSerArgThrProAlaArgArgArgSerAr 595
524 TGACCGCAACCCGACGACCGGACAAACGGCTCCGA...CCGTTTCCAC 570
595 gSerArgThrProThrArgArgSerArgSerArgThrProAlaArgA 612
571 AATCCGCGCTATGCTGACGCAAGAGTAGGAGGAGTATCAACGCGC 620
612 rGlyArgSerArgSerArgThrProAlaArgArgArgSerArgThrArg 628
621 CAACCGATACAGCCCGAGCTGACAGAGATCGCAATCGCGCGAGGCT 670
629 SerProValArgArgArgSerArgSerArgSerProAlaArgArgSerG 645
671 TCAA.....CGGCACTGCAGATATCGTCAAAA 698
645 yArgSerArgSerArgThrProAlaArgArgGlyArgSerArgSerArgT 662
699 ..CATCATCGCGCGCAGAGAAATTCGCG..... 728
```

---

```
662 hrProAlaArgArgGlyArgSerArgSerArgThrProAlaArgArgSer 678
729 ...CGACGCGATCCGTCGACGGGTATTAACGCAAGGCTCAAAACATTCG 775
679 GLyArgSerArgSerArgThrProAlaArgArgGlyArgSerArgSerAr 695
776 TCATGCAGCGCTTGGCTGCTTTCACCGCA.....AAACAAGATG 816
695 gThrProArgArgGlyArgSerArgSerArgSerLeuValArgArgGlyA 712
817 GCGCGCAT.....CAACGATTTGGCAGATATGCGCA..... 848
712 rSerHisSerArgThrProGlnArgArgGlyArgSerGlySerSerSer 728
849 .....ACTCAAAAGCTATGCCGAGC...AG 871
729 GLuArgLysAsnLysSerArgThrSerGlnArgArgSerArgSerAsnSe 745
872 CCATCCGCGCA.....T 882
745 rSerProGlnMetLysLysSerArgLysSerSerArgArgSerArgSerL 762
883 TGCGCAGTCCAAACCCCAATGCCGACAGCATGAGATGAGCCGTGACAA 932
762 eUserSerProArgSerLysAlaLysSerArg...LeuSerLeuArgArg 777
933 TATCTTTATGACAGCCAT.....CCCATCAAAAGGATTCGAGCTGCC 976
778 SerLeuSerGlySerSerProCysProLysGlnLysSerGlnThrProPr 794
977 GGGGAAATATAGGCTTGGCGGCATTCAC.....GGCACAT 1011
794 oArgArgSerArgSerGlySerSerGlnProLysAlaLysSerArgThr. 810
1012 CCTGTCAACCGGTGCGAGATGGCGCGCATCGCATTCGCCAAGGAAATC 1061
811 ProProArgArgSerArgSerSerSerSerProProLysGlnLysSe 827
1062 CGCCGTACAGCAATTTTGGCGATGCGGCATACGCCAAATCCCGTCCC 1111
827 rLysThrProSerArgGlnSerHisSerSerSerSerProHisPro... 842
1112 CTTACCATTCGCCAAATATCCGTTCAAACTTGACAGCAGCGTACGCAAA 1161
843 .....LysValLysSerGlyThrProProArgGlnGly... 853
1162 GAAACATCACCTCTCAACCGTCCGCGCTCAACCGGCAAAATGTCAA 1211
854 .....SerLeuThrSerProGlnAlaAsnGlnGlnSerValThr 866
1212 ACTGCGACAGCAACGCCCAACAGACAGCGCTACCGTTTGACGGTAAAG 1261
866 r.....ProGlnArgArgSerCysPheGlnSerSerP 877
1262 GGTTCGGAATTTGAGAAGCACGTGAATATGATACAGAGCTGATATT 1311
877 roAspProGlnLeuLysSerArgThrPro.....SerArg 888
1312 CAAGAATTAATCGGGGGGTATACCTAAGCTAAGCTGTGTTGAGAC 1361
889 HisSerCysSerGlySerSerProArgValLysSer.....SerTh 903
1362 GAACCGAGATGGGAGTTGATAGAACCTTAATTAATTCACAACCTGCTG 1411
903 rProProArgGlnSerProSerArgSer.....SerSerProg 916
1412 AGCAGGTGAGAAAATGTTCAAGAAACAGAGAGAGACAGT 1458
916 InProLysValLysAlaLeuLeuSerProArgGlnArgSerHisSer 931
```

seq\_name: sp\_human:Q90J35  
seq\_documentation\_block:

ID 090035 PRELIMINARY; PRT; 2752 AA.  
 AC 090035;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE RNA BINDING PROTEIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ohkaki S., Umeki K., Sawada Y.;  
 RT "Homo sapiens mRNA for RNA binding protein, complete cds."  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB016092; BAA83718.1;  
 DR InterPro: IPR002965; P\_Rich\_extensn.  
 DR PRINTS: PR01217; PRICHEXTENS.  
 SQ SEQUENCE 2752 AA; 299672 MW; 109C6AF181097123 CRC64;

alignment\_scores:  
 Quality: 157.00 Length: 617  
 Ratio: 0.584 Gaps: 30  
 Percent Similarity: 43.598 Percent Identity: 22.690

alignment\_block:  
 US-09-303-518D-463 x 090035

Align seg 1/1 to: 090035 from: 1 to: 2752

```

11 CCGCAAAATATCCCTTATTCGTGTCATCTGGCACTGTGCTGCGGATG 60
   ||||| |||||::: |||||
364 ProAlaProThrProLeuLeu.....AlaGluArgI 374
   ::::: |||||::: |||||
61 CATGCACAGCCCTCATGATTGGCAAGATCCCTTATCCGGCAGG...T 107
   ::::: |||||::: |||||
374 sglYgIserProGlnProLeuAlaThrThrProLeuSerGlnIuprov 391
   ::::: |||||::: |||||
108 TCTCGACCGCAGCATTTGCAACCCGAGGAAATACCC... 148
   ::::: |||||::: |||||
391 alasnProSerGluAlaSerProThrArgspArgSerProPolys 407
   ::::: |||||::: |||||
149 .....TATTCGACGACGAGGGGAGCTTGCCNAG 177
   |||||::: |||||
408 SerProGlnIuysIeuProGlnSerSerSerSerGluSerSerProse 424
   ::::: |||||::: |||||
178 CGCAACGCGCATATCGGATTGGAAACATACA..... 209
   ::::: |||||::: |||||
424 r.ProGlnProThrIysValSerArgHisAlaSerSerProGluSer 440
   ::::: |||||::: |||||
210 ...AAGCATCAGTTGGGCA..... 227
   |||||::: |||||
441 ProIysProAlaProAlaProGlySerHisArgGluIleSerSerSer 457
   ::::: |||||::: |||||
228 ....CCTGATGATTCAACAGCGCGCTTGAGAAATATCGGCTACAT. 272
   ::::: |||||::: |||||
457 oThSerIysAsnIysSerHisIsgIyArgAlaIalysArgAspIysSerHis 474
   ::::: |||||::: |||||
273 .....TGR 275
   ::::: |||||::: |||||
474 erHisIhrProSerArgArgMetGlyArgSerArgSerProAlaIhrAla 490
   ::::: |||||::: |||||
276 CCGCTTTTCGATCAGCGGCAAAATCCATTC..... 308
   |||||::: |||||
491 LysArgIglYArgSerArgSerArgThrProThrIysArgIlyHisSerAr 507
   ::::: |||||::: |||||
309 .....GCCCTTGACAAACCATGCCCTCACATTCGGATTGAC 345
   ::::: |||||::: |||||
507 gSerArgSerProGlnIhrpArgArgSerArgSerArgAlaGlnArgTrpAl 524
   ::::: |||||::: |||||
346 GAACCGGTATGCCGTTGACGATTCAGGCTTACCGCATTCATTCGGA 395
   |||||::: |||||

```

```

524 rYserArg.....SerProGlnArgArgGly 532
   ::::: |||||::: |||||
396 CGGATACGAAACACCATCCGC..... 416
   |||||::: |||||
533 ArgSerArgSerProGlnArgProGlyTrpSerArgSerArgAsnThrGl 549
   ::::: |||||::: |||||
417 .....CGACGGC 423
   ::::: |||||::: |||||
549 nArgArgIglYArgSerArgSerAlaArgArgGlyArgSerHisSerArgS 566
   ::::: |||||::: |||||
424 TATGACGGCCACAGCGCGGCTTATCCGCTCCCAAGCGCGAGGGA 473
   ::::: |||||::: |||||
566 erProAlaIhrArgGlyArgSerArgSerArgThrProAlaArgArgGly 582
   ::::: |||||::: |||||
474 TATTTACAGTACACATAAAGCGCTTGCCCAAAATATCCGCTCAACC 523
   ::::: |||||::: |||||
583 .....ArgSerArgSerArgThrProAlaArgArgSerAr 595
   ::::: |||||::: |||||
524 TGACCGAACCCGACGACCGGCAACGCGCTTGCCGA...CGGTTCCAC 570
   ::::: |||||::: |||||
595 gSerArgThrProThrArgArgArgSerArgSerArgThrProAlaArgA 612
   ::::: |||||::: |||||
571 AATGCCGCGCTATGCTGACGCAAGAGTAGCGGACGGATTCAAAGCGCG 620
   ::::: |||||::: |||||
612 rglYArgSerArgSerArgThrProAlaArgArgArgArgThrArg 628
   ::::: |||||::: |||||
621 CACCGGATACAGCCCGAGCTGACACAGATCGGCAATGCCCGGAAGCCT 670
   ::::: |||||::: |||||
629 SerProValArgArgArgSerArgSerArgSerProAlaArgArgSerGl 645
   ::::: |||||::: |||||
671 TCAA.....CGGCATGACAGATATGTCAAAAA. 698
   ::::: |||||::: |||||
645 YArgSerArgSerArgThrProAlaArgArgGlyArgSerArgSerArgT 662
   ::::: |||||::: |||||
699 ...CATCATCGCGCGGCGAGGAGAAATGTCGG..... 728
   ::::: |||||::: |||||
662 hrProAlaArgArgIglYArgSerArgSerArgThrProAlaArgArgSer 678
   ::::: |||||::: |||||
729 ...CGACGCGATGCCGTGACGGGTATTAAGCGAAGCTCAAAACATTGCTG 775
   ::::: |||||::: |||||
679 GlYArgSerArgSerArgThrProAlaArgArgIglYArgSerArgSerAr 695
   ::::: |||||::: |||||
776 TCATGCACAGGCTTGCTGCTCCACCGA.....AAACAAAGATG 816
   ::::: |||||::: |||||
695 gThrProIhrArgIglYArgSerArgSerArgSerLeuValArgArgGlyA 712
   ::::: |||||::: |||||
817 GCGCGCAT.....CAACGATTGGCAGATATGGCGCA..... 848
   ::::: |||||::: |||||
712 rYserHisSerArgThrProGlnArgArgGlyArgSerGlySerSerSer 728
   ::::: |||||::: |||||
849 .....ACTCAAGACTATGCCGACG...AG 871
   ::::: |||||::: |||||
729 GluArgIysAsnIysSerArgThrSerGlnArgArgSerArgSerAsnSe 745
   ::::: |||||::: |||||
872 CCATCCGGCA.....T 882
   ::::: |||||::: |||||
745 rSerProGlnMetIysIysSerArgIleSerSerArgArgSerArgSerL 762
   ::::: |||||::: |||||
883 TGGGAGTCCAAACCCCAATGCCGACAGGACATAGAACGCGTCAGCAA 932
   ::::: |||||::: |||||
762 eUserSerProArgSerIysAlaIysSerArg...LeuSerLeuArgArg 777
   ::::: |||||::: |||||
933 TATCTTTATGACGACAT....CCCATCAAGAGGATGGAGTGTC 976
   ::::: |||||::: |||||
778 SerLeuSerGlySerSerProCysProIysGlnIysSerGlnIhrPro 794
   ::::: |||||::: |||||
977 GGGGAAATACGCTTGCGCGCATCAC.....GGCACAAT 1011
   ::::: |||||::: |||||
794 oArgArgSerArgSerGlySerSerGlnProIysAlaIysSerArgThr 810
   ::::: |||||::: |||||
1012 CCTGTCAACGGTCCAGATGGCGCGCATCGATTCCGGAAGGAATTC 1061
   ::::: |||||::: |||||
811 ProProArgArgSerArgSerSerSerSerProProIysGlnIysSe 827
   ::::: |||||::: |||||

```

```

1062 CGCCGTCACGCAATTTGGCGATCGGCAATCCGCAATACCGGCC 1111
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
827 rlysthrProserArgInserHisSerSerSerProHisPro.... 842
1112 CTTACCATTCGCCGAATATCCGTTCAACTGGAGCAGCTTACGGCAAA 1161
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|
843 .....LysValLysSerGlyThrProProArgGlnGly... 853
1162 GAAACATACCTCTCTCAACCGTCGCCGCGCAAAAGCGCAAAATGTCAA 1211
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
854 .....SerIleThrSerProGlnAlaSerGlnGlnSerValThr 866
1212 ACTGGCAGACCAACGCCACCCGCAAGAGCGGTACCGTTGACCGTAAG 1261
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|
866 r.....ProGlnArgArgSerCysPheGlnSerP 877
1262 GGTTCGCAATTTTGAGAGCAGCTGAATATGTAGTACGAGCTGATATT 1311
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|
877 rAspProGlnLeuLysSerArgThrPro.....SerArg 888
1312 CAAGATATTCGGGGGGGGGTATACCTAAGGCTAAGCCTGTGTATGC 1361
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|
889 HisSerCysSerGlySerSerProProArgValLysSer.....SerThr 903
1362 GAAACCGGATGGAGGTTGATAGCAAGCTTAATTAATTGACAACTGCTG 1411
      : |||||:::|:::|:::|:::|:::|:::|:::|:::|
903 rProProArgGlnSerProSerArgSer.....SerSerProG 916
1412 AGCAGGTGAGCAAAATGTTTCAGAAACGAGACAGAGTCAAGACT 1458
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|
916 InProLysValLysAlaIleLeuSerProArgGlnArgSerHisSer 931
seq_name: sp_human:09UP82

```

```

seq_documentation_block:
ID 09UP82; PRELIMINARY; PRT; 944 AA.
AC 09UP82;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE MUCIN 4 (FRAGMENT).
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9077625; PubMed=9858486;
RA Gipson I.K., Spurr-Michaud S., Moccia R., Zhan Q., Toribara N.,
RA Ho S.B., Garibulo A.R., Hill J.A. III;
RT "MUC4 and MUC5B transcripts are the prevalent mucin messenger
RT ribonucleic acids of the human endocervix.",
RL Biol. Reprod. 60:58-64(1999).
DR EMBL; AF058803; AAC34750.1; -.
FT NON_TER 1
FT NON_TER 944
FT SEQUENCE 944 AA; 96314 MW; 5887CE9EC50855B5 CRC64;

```

```

alignment_scores:
  Quality: 155.50      Length: 500
  Ratio: 0.685        Gaps: 20
  Percent Similarity: 45.400  Percent Identity: 21.800

```

alignment\_block:

US-09-303-518D-463 x 09UP82

Align seg 1/1 to: 09UP82 from: 1 to: 944

```

9 TTCCGCAAAATATCCCTTATCTGTGCATACT..... 41
  ::::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

407 TrpProSerSerPheSerSerLysGlyHisThrThrTrpSerGlnThrG 423
42 .....GGCAGTGTGCTGCCATGCATGACACACCCCTCAGATTGGCAAC 87
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|
423 uLeuProSerThrSerThrGlyAlaIleThrArgLeuValThrGlyAsp 440
88 GATCCCTTTTCCGCGAGGTTCTCGACCGTC..... 118
      ||| |||||:::|:::|:::|:::|:::|:::|
440 roserThrGlyThrAlaGlyThrIleProArgValProSerLysValSer 456
119 AGCATTTGCAACCCGACGGGAATACACCTATTCGGCAGCGAGGGAG 168
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|
457 AlaIleGlyLysIleProGlyIleProThrThrThrSerSerHisSerThr 473
169 CTTGCCCAAGCGCAAG.....GCCATATCGATTGGGAACATACA 209
      |||||:::|:::|:::|:::|:::|:::|:::|:::|
473 rLeuProLysThrThrGlyAlaGlyAlaGlnThrGlnThrGlnGlnG 490
210 AACGCCATCGATTGGGCCACCTGATGATCAACAGCGCGCGCTTGAAGAA 259
      :::|:::|:::|:::|:::|:::|:::|:::|:::|
490 hngGlyThrThrGly..... 494
260 ATATCGGCTACATTTGTCCGCTTTTCCGATCAGCGGACAAATTCATCG 309
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|
495 ...GluAlaLeuLeuSerSerProSerGlySerValThrGlnMetIleGly 510
310 CCTTGACACACCATTCCTACATTCGATTCGATTCGACGAAG..... 349
      : ::::|:::|:::|:::|:::|:::|:::|:::|:::|
510 sThrAlaThrSerProSerSerSerProMetLeuAspArgHisThrSerG 527
350 .....CCGGTAGCCCGTTGACGGATTCAGCTTACCCCATTCAT 391
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|
527 InGlnIleThrThrAlaProSerThrAsnHisSerThrIleHisSerThr 543
392 GGGACGGATACGACACATCCCGCGGAGGCTATGACGGGACACAGGCG 441
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|
544 SerThrSerProGlnGlnSerProAlaValSerGlnArgGlyHisThrG 560
442 GCGCGGTATCCCGCTCCCAAGGCGCGAGGATATACAGCTGACAGAT 491
      ||| |||||:::|:::|:::|:::|:::|:::|
560 nAlaProGlnThrThrGlnGlnSerGln.....ThrThrA 572
492 AAAGGCGGTGCCCAAAATATCCGCTCAAC..... 523
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|
572 rSerValSerProMetThrAspThrLysThrValThrThrProGlySer 588
524 ..... 589 SerPheThrAlaSerGlyHisSerProSerGlnIleValProGlnAspAl 605
553 CTTGCCGACCGTTTCCACATGCGCGGCTATGCTGACGCAAGGAGTAGG 602
      : |||||:::|:::|:::|:::|:::|:::|
605 a...ProThrIleSer..... 610
603 CGACGATTTCAACGCGCCCGATACAGCCGCGAGCTGACAGATGCG 652
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|
610 lAlaIleThrThrPheAlaProAlaProThrGlyAspGlyHisThrThrGln 626
653 GCATGCGCGCGAAGCCTTCAAGGCGACATGAGATATGTCGCAAAACATC 702
      ||| |||||:::|:::|:::|:::|:::|:::|
627 Ala.....ProThrThrAlaLeuGlnIleValProSerSerHis 639
703 ATGCGCGCGGCGAGAGAAATGTGCGGCGCAGCGCATCGCGGAGGTAT 752
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|
639 sAspAlaThrLeuLysProSerGlyGlyThrSerLeu..... 651
753 AAGCGAGGCTCAAAACATTTGCTGCATGACAGCGCTTGCTGCTGCTTCA 802
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|
652 ..SerLysThrGlyAlaLeuThrLeuAlaSerValSerThrPro 667
803 .....CGAAACACAGATGCGCGCATCAACGATTTGGACGATATGCGG 846
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|
668 GlyLysProGlnGlyGlnThrThrSerAla..... 677

```

```

847 CAACTCAAGAACTATCCGCGAGCAGCCATCCGGAT..... 883
678 .....||| |||||:||||:|
678 Ser1aSer1ThrProAspThrAlaAlaIleMetThrIstHrIstG 694
884 ..... GGGCAGTCCAAACCCCAATGCCGCAC 910
694 1n1a1aSer1ThrGlu1aSer1Gly1nThr1nThrSer1uPro1a 710
911 AAGCA.....TAGAGCCGC 927
711 Ser5GlySerArgThrThrSer1aGlyTh1AlaThrProSerSerSe 727
928 AGCAATATCTTTATGGACGCATCCCATCAAGAGATTGAGACTGTCCG 977
727 rG1a1aSer1GlyTh1nThrProSer1GlySer1Gly1IleSerThnSerg 744
978 GGGAAATACGCGCTGGCGGCATCA.....CGGCACATCCGG 1015
744 1y1uTh1rTh1ArgPheSerSerAsnProSerArgAspSerH1ThrThr 760
1016 TCAAGCGGTCCGAGATGGCGCGATCCGATCCGACGAAGGAATCCGCC 1065
761 GlSer1ThrThrGlu1nLeu1aSer1aSer1aSer1aGly1a1a1leP 777
1066 GTCAAGCACAATTTGCCGATCGGCACATACGCCCAATACCCGCCCTTA 1115
777 oValSerThrGlyMetAlaSerSer1LeValProGlyThrPheAs1Prot 794
1116 CCAATCCCGAAATATCCGTT.....CAAACTTGACAGCCGTT 1153
794 hr1eSer1Glu1aSer1ThrAla1GArgProThrGly1nSerSerPro 810
1154 ACGCCAAAGAAACATCACCTCTCAACCGCGCGCGCAACG.... 1199
811 Th1rProSer1aSer1Pro1n1uTh1Ala1a1IleSerArgMeta 827
1200 .....CAAAATGTCAACTGGCAGA..... 1220
827 1aG1nThr1nArgThrArgThrSerArg1GlySerAspThr1IleSer1u 843
1221 ..... CCAAGCGCAGCC 1232
844 Alaserg1n1a1ThraspThrPheSerThrValProThrProPro 859

seq_name: sp_human:076037

seq_documentation_block:
ID: 076037 PRELIMINARY; PRT: 1045 AA.
AC 076037;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MUC1N PRECURSOR (FRAGMENT).
CN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON MUCOSA;
RA MEDLINE=98285542; PubMed=9620877;
RA NOLLET S., Moniaux N., Maury J.P., Petitprez D., Degand P., Laine A.,
RA Porchet N., Aubert J.P.;
RT "Human Mucin Gene MUC4: Organization of Its 5' Region and of Its
RT Polymorphic Central Tandem Repeat Array.";
RT Biochem. J. 332:739-748(1998).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON MUCOSA;
RA MEDLINE=20381033; PubMed=10920259;
RA Choudhury A., Moniaux N., Wimpenny J.P., Hollingsworth M.A.,

```

RA	Albert J.P.,	Batra S.K.;		
RT	"Human MUC4	matrix cDNA and its	variants in pancreatic carcinoma."	
RL	J. Biochem.	128:233-243(2000).		
DR	EMBL: A7000281;	CA03985.1;		
KW	Signal.			
FT	SIGNAL.	1	27	POTENTIAL.
FT	CHAIN	28	>1045	MUCIN.
FT	NONLEADER	1045	1045	
SQ	SEQUENCE	1045 AA;	106737 MW;	26112DF28E5D60A0 CRC64;

alignment_scores:	
quality:	155.50
ratio:	0.685
Percent Similarity:	45.400
	length:
	gaps:
	percent identity:
	500
	20
	21.800

alignment\_block:  
ms-00-203-F19D-463 x 076037

Align seg 1/1 to: 076037 from: 1 to: 1045

9	TTCCCGCAAAATATCCCTATTGCTGCATACG	41
10		
493	TRPProSerSerPheSerTyrGlyHisThrThrTrpSerGlnThrL	509
42	....GACATGTCCTGCCGATGCATGCACACGGCTCAGATTTGGCAAC	87
509	UleuProSerThrThrSerThrGlyAlaAlaThrThrGluLeuValThrGlyAsnP	526
88	GATCCCTTTATCCGGCAGGTTCTGCACCGTTC	118
526	roSerThrGlyAlaAlaGlyThrLleProArgValProSerTyrValSer	542
119	AGATTTGCACCCGACGGAGAAATACACCGCTATTGGCAGCAGGGGGAG	168
543	AlaIleGlyLysProGlyLysProThrThrTyrSerSerHisSerThr	559
169	CTTGGCCNACGGCAGACG	209
559	tleuProLysThrThrGlyAlaGlyAlaGlnThrGlnTrpThrGlnLys	576
210	AAGCATCAAGTTGGGCGACCTGATGATTCAACAGGCGCCGTTGAAGAA	259
576	hrGlyThrThrGly	580
260	ATATGGCGTACATGTCGGCTTTTCCATGCAGGACAAATTCATTCG	309
581	...GlnAlaLeuLeuSerProSerTyrIleValThrGlnMetIleY	596
310	CCCTGACACACCATCCTCACAATTCCGATTCTGACAGAG	349
596	strAlaThrThrProSerSerProMetLeuAsnArgHisThrSerG	613
350	.....CCGGTAGTCCCGTTGACGGATTCAGCCCTTACCAGTCAT	391
613	InglnIleThrThrAlaProSerThrAsnHisSerThrIleHisSerThr	629
392	GGGACGATGCAGAACACATCCCGCGGACGGTATGACGGCCAGACGGC	441
630	SetThrSerProGlnSerProAlaValSerGlnArgGlyHisThrG	646
442	GGCGGCTATCCCGCTCCCAAGAGCGGAGGATATATGACAGTCAGCAT	491
646	mlaProGlnThrThrGlnGlnSerGln	658
492	AAAAGCGGTGCCCAAAATATCCGCTCAAC	523
658	rgSerValSerProMetThrAspThrLysThrValThrProGlySer	674
554	.....TGACCGAACACCGACGACGGACGAC	552
675	serPheThrLaseArgLysThrSerProSerGlnLysValProGlnAspAl	691



```

553 CTGCGACGCGTTTCCACATGCGGCGCATGCTGACGCAAGAGTAGG 602
691 : |||||: |||||
691 A...Prothriliser.....A 696
603 CGAGCGATTCAAAGCGCCACCCATACAGCCCGAGCTGACAGATCGG 652
696 : |||||: |||||
696 laaIatThrPheAlaProAlaProThrGlyAspGlyHisThrGln 712
653 GCAATGCGCGCGAGCCCTCAAGCGCATGACATGTCGCAAAACATC 702
713 Ala.....Prothriliser.....A 713
703 ATCGCGCGCGAGAGAAATGTCGCGCGAGCGATGCGCGAGGAT 752
725 : |||||: |||||
725 sasAlaThrLeuGlyProSerGlyGlyThrLeu..... 737
753 AAGGAGGCTCAACATGCTGTCATGACAGCGCTTGCTGCTTCCA 802
738 : |||||: |||||
738 SerLysThrGlyAlaLeuThrLeuAlaAsnSerValSerThrPro 753
803 : |||||: |||||
803 : |||||: |||||
754 GlyGlyProGlyGlyGlyThrSerAla..... 763
847 CAATCAAGACTATGCGCGAGCGATGCGCGAT..... 883
764 : |||||: |||||
764 SerAlaSerThrSerProAspThrAlaAlaMetThrHisG 780
884 : |||||: |||||
884 : |||||: |||||
780 lnaIaGlySerThrGlyAlaSerGlyGlyThrGlyThrSerGlyProAla 796
911 AAGCA.....TAGAAGCGTC 927
797 SerSerGlySerAlaGlyThrSerAlaGlyThrAlaThrProSerGly 813
928 AGCAATCTTTATGCGCGATGCGCGATGCGCGATGCGCGATGCGG 977
813 : |||||: |||||
813 rGlyLysThrGlyThrThrProSerGlySerGlyGlyThrSerG 830
978 GCGAAATATGCGCTGCGCGATGCGCGATGCGCGATGCGCGATGCGG 1015
830 : |||||: |||||
830 LysGlyThrThrArgPheSerSerAspProSerAlaGlySerHisThr 846
1016 TCAAGCGCTGCGAGATGCGCGATGCGCGATGCGCGATGCGCGATGCGG 1065
847 GlnSerThrThrGlyLeuSerAlaSerAlaSerHisGlyAlaLeu 863
1066 GTGAGCGCAATTTGCGCGATGCGCGATGCGCGATGCGCGATGCGG 1115
863 : |||||: |||||
863 oValSerThrGlyMetAlaSerSerLeuValProGlyThrPheHisPro 880
1116 CCAATGCGCGATGCGCGATGCGCGATGCGCGATGCGCGATGCGG 1153
880 : |||||: |||||
880 hLeuSerGlyAlaSerThrAlaGlyArgProThrGlyGlnSerSerPro 896
1154 AGCGCAAAAGAAATCACCCTCTCAGCGCTGCGCGATGCGCGATGCGG 1199
897 : |||||: |||||
897 ThrSerProSerAlaSerProGlnGlyThrAlaAlaLeuSerArgMeta 913
1200 : |||||: |||||
1200 : |||||: |||||
913 lalGlnThrGlnArgThrArgThrSerArgGlySerAspThrLeuSerLeu 929
1221 : |||||: |||||
1221 : |||||: |||||
930 AlaSerGlnAlaThrAspThrPheSerThrValProProThrProPro 945
seq_name: sp_human:Q9H4D6
seq_documentation_block:
ID Q9H4D6 PRELIMINARY; PRT; 1107 AA.
AC Q9H4D6;

```

```

DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE MUC4 PROTEIN SPLICED VARIANT SV11.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Choudhury A., Moniaux N., Ringel J., Aubert J.P., Batra S.K.;
RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ277505; CAC14585.1;
SQ SEQUENCE 1107 AA; 113576 MW; 2669ADBFD5A53721 CRC64;

alignment_scores:
Quality: 155.50 Length: 517
Ratio: 0.713 Gaps: 21
Percent similarity: 42.166 Percent identity: 22.631

alignment_block:
us-09-303-518d-463 x Q9H4D6 ..

Align seg 1/1 to: Q9H4D6 from: 1 to: 1107

9 TTCGCGCAAAATATCCCTATCTGTCATCT..... 41
493 TrpProSerSerPheSerSerLysGlyHisThrTrpSerGlnThrGly 509
42 : |||||: |||||
42 : |||||: |||||
509 uLeuProSerThrSerThrGlyAlaAlaThrArgLeuValThrGlyAsp 526
88 GATCCCTTATCCGCGAGTTCGCGATGCGCGATGCGCGATGCGCGATGCGG 118
526 roSerThrArgAlaAlaGlyThrLeuProAlaGlyAlaProSerLysValSer 542
119 AGCATTTGCAACCCGCGGAAATACACCTATTCGCGAGCGAGGAGGAG 168
543 AlaLeuGlyLysProGlyLysProThrThrLysSerSerHisSerThrTh 559
169 CTGCGCGCGCGAGCG.....GCCATATCGGATTCG..... 199
559 rLeuProLysThrThrGlyAlaGlyAlaGlnThrGlnThrGlnGlnGly 576
200 : |||||: |||||
200 : |||||: |||||
576 hngLyThrThrGlyAlaLeuLeuSerSerProSerThrSerValIle 592
209 AAAGCATCAGTTGGCGCCACCTGATGATTCACAGGCGCGCTTG...AA 255
593 GlnMetLeuLysThrAlaThrSerProSerSerSerProMetLeuAspAr 609
256 GGAATATCGCGCTACATTTGCGCGCTTTCGATACGCGGCGCAAAATCCA 305
609 gHisThrSerGlnGlnThrThrAlaProSerThrAsnHisSerThrTr 626
306 TTGCGCGCTGCGACACATGCGCT...CACATTCGATTCG..... 343
626 lHisSerThrSerThrSerProGlnLysSerProAlaValSerGlnArg 642
344 : |||||: |||||
344 : |||||: |||||
643 GlyHisThrArgAlaProGlnThrThrGlnLysSerGlnThrThrArgse 659
352 GGTAGTCCCGTTGACGATTCAGCCTTACCGCATTCATGGGAGGAGATA 401
659 rValSerProMetThrAsp.....ThrLysT 668
402 CGAACACCATCCCGCGAGCGGCTATGACG.....GCCACAGGCGCGCG 445

```

```

668 hVvalThrThrProGlySerPheThrAlaSerGlyHisSerProSer 684
446 GCTATCCCGCTCCCAAGAGCGGAGGATATATACAGCTACAGATATAA 495
685 GluIleValProGlnAspAla..... 691
496 GGCCTTGGCCCAAAATATCGCCTCACTGACCGCAACCCGACGCGG 545
692 .....ProThrIleSerAlaIleThrThrPheAlaProAlaProThrG 706
546 ACAAGCGCTTGGCGACCGCTTCCACATCGCGCTTATGCTGACGCAAG 595
706 LysMetIleHisThrThrGlnAlaProThrThrAlaLeu..... 718
596 GAGTAGCGCAGCATTCAAACGCGCCACCGATACAGCCGCGAGCTGAC 645
719 .....GlnAlaAlaProSerSerH 725
646 AGATCGGCGAATGCCCGCAACCTTCACAGCGCATGCAAGATATGCTGA 695
725 sAspAlaThrLeuGlyProSerGlyGlyThrSerLeu.....SerL 739
696 AACCATCATTCGCGCGCAGAGAAATGTCCGCGCAGGCGATGCCGTGC 745
739 ysthrGly..... 741
746 AGGGTATTAACGAGAGGCTCAAAACATGCTGTATGACAGCGCTGGGTCTG 795
742 .....AlaLeuThrIleuAlaAsnSerValValSerThrPr 753
796 CTTTCACCGGAAACAGAGATGGCGCATCAACGATTTGGCATATATGAC 845
753 oglyGlyProGlyGlnThrThrSerAla..... 763
846 GCAACTCAAAAGATATCGCGCAGCAGCAGCATCCGCGATT..... 883
764 ..SerIaSerThrSerProAspThrAlaAlaIleMetThrHisThrHis 779
884 .....GGGCGATCCCAAAACCCCAATGCCGA 909
780 GlnAlaGlnSerThrGlnAlaSerGlyGlnThrGlnThrSerGlnProAl 796
910 CAAGGCA.....TAGAAGCCGT 926
796 aserSerGlySerArgThrThrSerAlaGlyThrAlaThrProSerSerS 813
927 CAGCAATATCTTATGAGCCATCCCATCAAAAGGATGAGCTGTCC 976
813 erglyAlaSerGlyThrThrProSerGlySerGlnGlyIleSerThrSer 829
977 GGGGAAATACGGCTTGGGCGGATCA.....CGGACATCT 1014
830 GlyGlnThrThrArgPheSerSerAsnProSerArgAspSerHisThrH 846
1015 GTCAAGCGGTCCGAGATGGCGGATCGATTCGCAAAAGGCAAAATCCGC 1064
846 rGlnSerThrThrGlnLeuLeuSerAlaSerHisGlyAlaIleP 863
1065 CGTCAGCGCAATTTTGGCGATCGGCGATAGCCCAAAATCCGCTCCCTT 1114
863 rovalSerThrGlyMetAlaSerSerIleValProGlyThrPheHisPro 879
1115 ACCTTCGCGAAATATCGCT.....CAAACTGGAGCGAGCT 1152
880 ThrIleuSerGlnAlaSerThrAlaGlyArgProThrGlyGlnSerSerP 896
1153 TACGGCAAAAGAAACATCACCTCTCAACCGCTCCGCGCTCAACG... 1199
896 ofThrSerProSerAlaSerProGlnGlu..ThrAlaAlaIleSerArgMet 912
1200 .....CAAAATATCTCAAACTGGCAGA..... 1220

```

```

913 AlaGlnThrGlnArgThrGlyThrSerArgGlySerAspThrIleSerIe 929
1221 .....CCACGCGCACCC 1232
929 uAlaSerGlnAlaThrAspThrPheSerThrValProThrProPro 945
seq_name: sp_human:Q96ZM2

```

```

seq_documentation block:
ID Q96ZM2 PRELIMINARY; PRT: 1157 AA.

```

```

AC Q96ZM2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE MUC4 PROTEIN SPLICED VARIANT SV13 (MUC4 PROTEIN SPLICED VARIANT
DE SV12).
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREATIC TUMOUR.
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ400850; CAC14135.1; -
DR EMBL: AJ400849; CAC14134.1; -
SQ SEQUENCE 1157 AA: 118925 MW; F46F039320969A9C CRC64;

```

```

alignment_scores:
Quality: 155.50 Length: 517
Ratio: 0.713 Gaps: 21
Percent Similarity: 42.166 Percent Identity: 22.631

```

```

alignment_block:
US-09-303-518D-463 x Q96ZM2 ..
Align seg 1/1 to: Q96ZM2 from: 1 to: 1157

```

```

9 TTCGCGCAAAATATCCCTTATCTGTCTCATCT..... 41
493 TrpProSerSerPheSerSerGlyHisThrThrTrpSerGlnThrG 509
42 .....GGCAGTGCGCTGCGCATGATCGACAGCGCTCAGATTGGCAAC 87
509 uLeuProSerThrSerThrGlyAlaIleThrArgLeuValThrGlyAsn 526
88 GATCCCTTATCGCGCAGGTTCTGACCGTC..... 118
526 roserThrArgAlaIleGlyThrIleProArgValProSerIleValSer 542
119 AGCATTTTGAACCGCAGCGGAAATACCACTATTCGAGAGAGGGGAG 168
543 AlaIleGlyGlnProGlyGlnProThrThrTySerSerHisThrH 559
169 CTTCGCGNAGCGCAAG.....GCCATATCGGATTGG..... 199
559 rIleuProIysThrThrGlyAlaGlyAlaGlnThrGlnThrPheGlnIu 576
200 .....GAACATAC 208
576 hrGlyThrThrGlyGlnAlaLeuLeuSerSerProSerTySerValIle 592
209 AAAGCATAGTGTGGCGCCATGATGATTCAAACAGCGCGCGTTG...AA 255
593 GlnMetIleIysThrAlaThrSerProSerSerSerProMetIleAsp 609
256 GGAATATCGGCTACATGTCGCTTTTCGATCAGGGGCAAAATTCGA 305
609 gHisThrSerGlnGlnIleThrThrAlaProSerThrAsnHisSerThrI 626

```





```

9 TTCCCGAAATATCCCTATTCTGTCCATCT..... 41
493 TrpProSerPheSerSerSerGlyHisThrThrTrpSerGlnThrG1 509
42 ...GGCAGTGTCCCTCCGATGCATGCACAGCCCTCAGATTGGCAAC 87
509 uLeuProSerThrSerThrGlnAlaAlaThrArgLeuValThrGlyAsn 526
88 GATCCCTTATCCGGCAGGTCTGTGACCGTC..... 118
536 roSerThrArgAlaAlaGlyThrIleProArgValProSerGlyValSer 542
119 AGCATTTTGAACCCGACGCGGAATATCACCATTTCGGAGACAGCGGGAG 168
543 AlaIleGlyGluProGlyGluProThrThrTyrSerSerHisSerThr 559
169 CTTCGCNAGCCGCAACG.....GCCATATCGGATTGG..... 199
559 rLeuProLysThrThrGlyAlaGlyAlaGlnThrGlnThrGlnGluT 576
200 .....GAAACAATAC 208
576 hrGlyThrThrGlyGluAlaLeuLeuSerSerProSerTyrSerValIle 592
209 AAAGCCATCAGTTGGCCACCTGATGATTCACAGCGCGCGTTG...AA 255
593 GlnMetIleLysThrAlaThrSerProSerSerSerProMetLeuAsp 609
256 GGAATATCGGCTACATTCGCTTTCCGATCAGCGGACAAATTCGA 305
609 gHisThrSerGlnIleThrThrAlaProSerThrAsnHisSerThrI 626
306 TTCGCCCTTGACAAACATGCT...CACATTCGATTCTG..... 343
626 leHisSerThrSerThrSerProGlnIleSerProAlaValSerGlnArg 642
344 .....ACGAAGCC 351
643 GlyHisThrArgAlaProGlnThrThrGlnIleSerGlnThrThrArg 659
352 GGTATGCCGCTTGACGATTCAGCTTTACCGCATTCATGGACGATGA 401
659 rValSerProMetThrAsp.....ThrLysT 668
402 CGAACACCATCCCGCCGCGCGCTATGACG.....GGCCACAGGCGGCG 445
668 hrValThrThrProGlySerSerPheThrAlaSerGlyHisSerPro 684
446 GCTATCCCGCTCCCAAGCGCGAGGATATATACAGCTACGACATPAAA 495
685 GluIleValProGlnAspAla..... 691
496 GCGGTTCCCAAAATATCCGCTCAACCTGACGCGACACCGCAGCAGG 545
692 .....ProThrIleSerAlaAlaThrThrPheAlaProAlaProThrG 706
546 ACAAGCGCTTGCCGACCGTTTCCCAATGCGCGCTATGCTGACGCAAG 595
706 LysAsnGlyHisThrThrGlnAlaProThrThrAlaLeu..... 718
596 GAGTAGCGCGAGGATTCAAAGCGCGACCGGATACGCGCGAGCTGAG 645
719 .....GlnAlaAlaProSerSerHis 725
646 AGATGGGCAATGCGCGCAAGCCTTCAACGCGACGACGAGATATGCTCA 695
725 sAspAlaThrLeuGlyProSerGlyGlyThrSerLeu.....SerL 739
696 AAACATCATCGCGCGCGAGGAATATCTCGCGCAGCGCATGCCGCTGC 745
739 ySerThrGly..... 741

```

```

746 AGGGTATTAAGCGAAGCCTCAACATTGCTGTCTATGCACGCGCTTGCTG 795
742 .....AlaLeuThrIleAlaAlaAsnSerValValSerThrPr 753
796 CTTTCCACCGCAAAACAGATGGCGGCATCCACGATTTGGCAGATGTGC 845
753 ogGlyGlyProGlyGlyGlnThrPheSerAla..... 763
846 GCAACTCAAAAGACTATCCGCGACAGCAGCATCCGCGATT..... 883
764 ..SerAlaSerThrSerProAspThrAlaAlaAlaMetThrHisThr 779
884 .....GGCAGTCCAAATCCCAATGCCGA 909
780 GlnAlaGluSerThrGlnAlaSerGlyGlnThrGlnThrSerGluProAl 796
910 CAAGCA.....TAGAAGCGCT 926
796 aserSerGlySerArgThrThrSerAlaGlyThrAlaThrProSerSer 813
927 CAGCAATATCTTTATGCGACGCCATCCCATCAAGAGATTGAGCTGCC 976
813 erGlyAlaSerGlyThrThrProSerGlySerGlyGlyIleSerThrSer 829
977 GGGGAAATATACGCTTGGGCGGCATCA.....CGGCACATCCT 1014
830 GlyGluThrThrArgPheSerSerSerAsnProSerArgAspSerHisThr 846
1015 GTCAACGCGTCCGACATGGCGCGCATTCGATCCGGAAGGAAATCCGC 1064
846 rGlnSerThrThrGlnLeuLeuSerAlaSerAlaSerHisGlyAlaIle 863
1065 CGTCACGCGCAATTTTCCGATGCGGATACGCCAAATACCCGCTCCCTT 1114
863 rValSerThrGlyMetAlaSerSerIleValProGlyThrPheHisPro 879
1115 ACCATTCCGAAATATCCGCT.....CAACTTGAGCAGCAGGT 1152
880 ThrLeuSerGluAlaSerThrThrAlaGlyArgProThrGlyGlnSerSer 896
1153 TACGCGAAAGAAACATCATCTCTCAACCGTCCGCGCTCAACGCG... 1199
896 oThrSerProSerAlaSerProGlnGluThrAlaAlaIleSerArgMet 912
1200 .....CAAAATGTCAAACTGGCAGA..... 1220
913 AlaGlnThrGlnArgThrGlyThrSerArgGlySerAspThrIleSerLe 929
1221 .....CCACGCGCCAGCC 1232
929 uAlaSerGlnAlaThrAspThrPheSerThrValProProThrProPro 945

seq_name: sp_human:09H486
seq_documentation_block:
ID 09H486 PRELIMINARY; PRT; 1201 AA.
AC 09H486:
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MUC4 PROTEIN SPLICER VARIANT SV16.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREATIC TUMOUR.
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ400853; CAC14138.1; -

```



GN MUC4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-PANCREATIC TUMOUR;  
 RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;  
 RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma";  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ400854; CAC14139.1;  
 SO SEQUENCE 1214 AA; 125245 MW; B89D6A84AD4A056 CRC64;

alignment\_scores:  
 Quality: 155.50 Length: 517  
 Ratio: 0.713 Gaps: 21  
 Percent Similarity: 42.166 Percent Identity: 22.631

## alignment\_block:

US-09-303-518D-463 x Q9H485

Align seg 1/1 to: Q9H485 from: 1 to: 1214

```

9  TTCCCGCAAAATATCCCTTATCTGTCCACT..... 41
   ::::::::::::::: ::::|
493  TrpProSerSerPheSerSerLysGlyHisThrTrpSerGlnThrG1 509
   ::::::::::::::: ::::|
42  ...GGCAGTGTGCTCCGATGCATGCACAGCGCTCAGATTGGCAAC 87
   ::::| ::::| ::::| ::::| ::::| ::::|
509  uLeuProSerThrSerThrGlyAlaAlaThrArgLeuValThrGlyLysn 526
   ::::| ::::| ::::| ::::| ::::| ::::|
88  GATCCCTTATCCGAGGTTCCTGACCGTC..... 118
   ||| ||||| ::::|
526  roserThrArgAlaAlaGlyThrIleProArgValProSerLysValSer 542
   ::::| ::::| ::::| ::::| ::::| ::::|
119  AGCATTTGCAACCCGACGGAAATACACCTATTCGACAGCAGGCGGAG 168
   ::::| ::::| ::::| ::::| ::::| ::::|
543  AlaIleGlyLysProGlyGluProThrThrLysSerHisSerThr 559
   ::::| ::::| ::::| ::::| ::::| ::::|
169  CTGCGCAAGCGCAAG.....GCCATTCGATTG..... 199
   ::::| ::::| ::::| ::::| ::::| ::::|
559  rLeuProLysThrThrGlyAlaGlyAlaGlnThrGlnThrGlnLys 576
   ::::| ::::| ::::| ::::| ::::| ::::|
200  .....GAAACATAC 208
   ::::|
576  hrGlyThrThrGlyAlaLeuLeuSerSerProSerTyrSerValIle 592
   ::::| ::::| ::::| ::::| ::::| ::::|
209  AAAGCATCATGTTGGCCACCTGATTCACAGCGCGCGTTG...AA 255
   ::::| ::::| ::::| ::::| ::::| ::::|
593  GlnIleLysThrAlaThrSerProSerSerSerProMetLeuAsp 609
   ::::| ::::| ::::| ::::| ::::| ::::|
256  GGAATATCGGCTACATTCGCTTCCGATTCGACAGCGGCAAAATCCA 305
   ::::| ::::| ::::| ::::| ::::| ::::|
609  gHisThrSerGlnIleThrThrAlaProSerThrAsnHisSerThr 626
   ::::| ::::| ::::| ::::| ::::| ::::|
306  TTCGCCCTTCGACACCATGCT...CACATTCGATTGTG..... 343
   ::::| ::::| ::::| ::::| ::::| ::::|
626  leHisSerThrSerThrSerProGlnLysSerProAlaValSerGlnArg 642
   ::::| ::::| ::::| ::::| ::::| ::::|
344  .....ACGAAGCC 351
   ::::|
643  GlyHisThrArgAlaProGlnThrThrGlnLysSerGlnThrThrArg 659
   ::::| ::::| ::::| ::::| ::::| ::::|
352  GGTAGTCCGTTGACGATTCACGCTTACGCGATCATTCGGACGGATA 401
   ::::| ::::| ::::| ::::| ::::| ::::|
659  rValSerProMetThrAsp.....ThrLys 668
   ::::| ::::| ::::| ::::| ::::| ::::|
402  CGAACACCATCCGCGCGACGGCTATGACG.....GGCACAGCGCGCG 445
   ::::| ::::| ::::| ::::| ::::| ::::|
668  hrValThrThrProGlySerSerPheThrAlaSerGlyHisSerProSer 684
   ::::| ::::| ::::| ::::| ::::| ::::|

```

```

446  GCTATCCCGCTCCCAAGCGCGAGGGATATATACAGCTACGACATMAAA 495
   |||
685  GlnIleValProGlnAspAla..... 691
   ::::|
496  GGCCTGCCCAAAATATCCGCTTCACCTGACCGGACAGACCGGACCGG 545
   ||| ||||| ::::|
692  .....ProThrIleSerAlaAlaThrThrPheAlaProAlaProThrG 706
   ::::| ::::| ::::| ::::| ::::| ::::|
546  ACAACGGCTTCCGACCGCTTCCCAATGCGCGCTATGCTGACGCAAG 595
   ||||| ::::|
706  LysnGlyHisThrThrGlnAlaProThrThrAlaLeu..... 718
   ::::| ::::| ::::| ::::| ::::| ::::|
596  GAGTAGCGGACGATTCAAACGCGCCACCGCATACAGCCCGGCTGAC 645
   ::::| ::::| ::::| ::::| ::::| ::::|
719  .....GlnAlaAlaProSerSerH 725
   ::::|
646  AGATCGGCAATGCGCGCGAAGCCTTCACGCGCATGCAATGCTGCAA 695
   ||| ::::| ::::| ::::| ::::|
725  sAspAlaThrLeuGlyProSerGlyGlyThrSerLeu.....SerL 739
   ::::| ::::| ::::| ::::| ::::| ::::|
696  AAACATCATCGCGCGCGACGAGAAATGTGCGCGCAGCGGATGCCGTG 745
   ||||| ::::|
739  ySThrGly..... 741
   ::::|
746  AGGTATTAAGCGAAGCTCAACATTCGTCTCATGCGCGCTTGGGTG 795
   ||| ||||| ::::|
742  .....AlaLeuThrLeuAlaAsnSerValValSerThrPr 753
   ::::|
796  CTTTCACGCAAAACAAAGATGGCGCGCATCAACGATTGGAGATATGCG 845
   ||||| ::::|
753  oGlyLysProGlyGlnThrThrSerAla..... 763
   ::::|
846  GCACATCAAGACTATCCGCGACGACCATCCGCGATT..... 883
   ::::|
764  ...SerAlaSerThrSerProAspThrAlaAlaAlaMetThrHis 779
   ::::|
884  .....GGCGAGTCCCAAAACCCCATGCGCGA 909
   ::::|
780  GlnAlaGlnSerThrGlnAlaSerGlyGlnThrGlnThrSerGlnProAl 796
   ::::|
910  CAAGGCA.....TAGAAGCGT 926
   ::::|
796  aserSerGlySerArgThrThrSerAlaGlyThrAlaThrProSerSer 813
   ::::|
927  CAGCAATATCTTATGCGACGCCATCCCATCAAAAGGATTCGACGTGTC 976
   ::::|
813  erGlyAlaSerGlyThrThrProSerGlySerGlyLysThrSer 829
   ::::|
977  GGGGAAATATACGCTTGGCGGCGCATCA.....CGGCACATCCT 1014
   ||||| ::::|
830  GlyLysThrThrArgPheSerSerSerProSerArgAspSerHisThr 846
   ::::|
1015  GTCACGCGTCCAGATGGCGCGCATTCGCAATTCGCGAAAGGGAATCCGC 1064
   ::::|
846  rGlnSerThrThrGlnLeuLeuSerAlaSerAlaSerHisGlyAlaIle 863
   ::::|
1065  CGTCAGCGCAATTTGCGGATGCGCGCATACGCCCAATACCGCTCCCTT 1114
   ::::|
863  roValSerThrGlyMetAlaSerSerIleValProGlyThrPheHisPro 879
   ::::|
1115  ACCATTCGCAAAATATCCGTT.....CAACCTTGAGCAGCGCT 1152
   |||||
880  ThrLeuSerGlyAlaSerThrAlaGlyArgProThrGlyGlnSerSerPr 896
   ::::|
1153  TACGGCAAAAGAAACATACCTCTCAACCGTGGCGCGCTCAACAGG... 1199
   ::::|
896  oThrSerProSerAlaSerProGlnGluThrAlaAlaIleSerArgmet 912
   ::::|
1200  .....CAAAATGTCAAACTGGCAGA..... 1220
   ::::|
913  AlaGlnThrGlnArgThrGlyThrSerArgLysAspThrIleSerLe 929
   ::::|
1221  .....CCAAGCGCACCC 1232
   ::::|

```

929 uAlaSerGlnAlaThrAspThrPheSerThrValProProThrProPro 945

seq\_name: sp\_human:Q9H484

seq\_documentation\_block: PRELIMINARY; PRT: 1215 AA.

ID Q9H484  
AC Q9H484  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, last annotation update)  
DE MUC4 PROTEIN SPLICED VARIANT SV18.  
GN MUC4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PANCREATIC TUMOUR;  
RA Choudhury A., Moriaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;  
RT Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ40085; CAC14140.1;  
DR InterPro: IPR003886; Nidogen\_ext.  
DR SMART: SM00539; NIDO; 1.  
SQ SEQUENCE 1215 AA; 125435 MW; 2B950CD46945D60D CRC64;

alignment\_scores:  
Quality: 155.50 Length: 517  
Ratio: 0.713 Gaps: 21  
Percent Similarity: 42.166 Percent Identity: 22.631

alignment\_block:  
US-09-303-518D-463 x Q9H484 ..

Align seg 1/1 to: Q9H484 from: 1 to: 1215

9 TTCCCGCAAAATATCCCTTATCTGTCATACAT ..... 41  
493 TTPProSerSerPheSerSerLysGlnHisThrThrTrpSerGlnThrGln 509  
42 ....GSCAGTGTGCTCCGATGCATGCACACGCTCAGATTGGCAAC 87  
509 uLeuProSerThrSerThrGlnAlaAlaThrArgLeuValThrGlyAsn 526  
88 GATCCCTTATCCGCGAGTCTCGACCCGTC ..... 118  
526 roSerThrArgAlaAlaGlyThrIleProArgValProSerLysValSer 542  
119 AGCATTTTCGAACCGGAGGGAATACCACTATTCGGCAGCAGGGGAG 168  
543 AlAlaGlyGlnProGlyGlnProThrThrLysSerSerHisSerThr 559  
169 CTTCGCCNAGGCGAAG.....GCCATATCGGATTG..... 199  
559 rLeuProLysThrThrGlyAlaGlyAlaGlnThrGlnTrpThrGlnGln 576  
200 .....GAAACATAC 208  
576 hrGlyThrThrGlyAlaAlaLeuLeuSerSerProSerTyrSerValIle 592  
209 AAAGCCATAGTTGGCCACCTGATTCATACAGCGGCGCTTG...AA 255  
593 GlnMetIleLysThrAlaThrSerProSerSerSerProMetLeuAsp 609  
256 GGAATATCGGCTACATGTCGCTTTCGATCAGCGGCAAAATTCGA 305  
609 gHisThrSerGlnGlnIleThrThrAlaProSerThrAsnHisSerThr 626  
306 TTCGCCCTTGCAACCAATGCT...CACATTCGATTCG..... 343

626 leHisSerThrSerThrSerProGlnGlnSerProAlaValSerGlnArg 642  
344 .....ACGAAGCC 351  
643 GlyHisThrArgAlaProGlnThrThrGlnGlnSerGlnThrThrArg 659  
352 GGTATGCCCTTGACGATTCAGCCTTTACCGCATTCATGGGACGATA 401  
659 rValSerProMetThrAsp.....ThrLysT 668  
402 CGAACACCATCCCGCGCGGCTATGACG.....GGCCACAGGCGGCG 445  
668 hrValThrThrProGlySerSerPheThrAlaSerGlyHisSerPro 684  
446 GCTATCCCGCTCCCAAGCGCGGAGGATATACAGCTACGACATAAA 495  
685 GluIleValProGlnAspAla..... 691  
496 GCGCTTGCCCAAAATATCCGCTTCACCTGACGACACGCGACCGG 545  
692 .....ProThrIleSerAlaAlaThrThrPheAlaProAlaProThr 706  
546 ACAAGGCTTCGCGACGCTTTCACAATGCCGCGCTATGCTGACGCAAG 595  
706 LysGlnGlyHisThrThrGlnAlaProThrThrAlaLeu..... 718  
596 GAGTAGCGGACGATTCAAACGCCGACCGCATACAGCCCGGAGCTGAC 645  
719 .....GlnAlaAlaProSerSerH 725  
646 AGATCGGCAATGCGCGGAGCGCTTCACGCGCACCTCAGATATGTCGA 695  
725 sasPrAlaThrLeuGlyProSerGlyGlyThrSerLeu.....SerL 739  
696 AAACATCATCGCGCGGCGGAGCAAAATGTGCGGCGGCGGATCCGTC 745  
739 ystHrGly..... 741  
746 AGGTTATAGCGGAAGGCTCAACATGTGTCATGACAGCGGCTGGGTG 795  
742 .....AlaLeuThrLeuAlaAsnSerValValSerThrPr 753  
796 CTTCACCGCAAAACAGATGCGCGCATCAACGATTTGGCAGATATGCG 845  
753 ogLysLysProGlnGlnGlnTrpThrSerAla..... 763  
846 GCAACTCAAAAGCTATGCGCGACAGCCATCCGCGAT..... 883  
764 ..SerAlaSerThrSerProAspThrAlaAlaAlaMetHisThrHis 779  
884 .....GGGAGTCGCAAAACCCCAATGCCGCA 909  
780 GlnAlaGlnSerThrGlnAlaSerLysGlnThrGlnThrSerGlnProAl 796  
910 CAAGGCA.....TAGAGCGGT 926  
796 aserSerGlySerArgThrThrSerAlaGlyThrAlaThrProSerSer 813  
927 CAGCAATATCTTTATGAGCAGCCATCCCATCAAGGATTTGGAGTGGC 976  
813 erGlyAlaSerGlyThrThrProSerGlySerGlyLysThrThrSer 829  
977 GGGAAATACGCGCTTGGGCGGATCA.....CGGCACATCCT 1014  
830 GlyGlnThrThrArgPheSerSerAsnProSerArgAspSerHisThr 846  
1015 GTCAAGCGGCTCGCAGATGCGCGCATTCGATTCGCAAGGGAATTCGCG 1064  
846 rGlnSerThrThrGlyLeuLeuSerAlaSerAlaSerHisGlyAlaIle 863  
1065 CGTCAGCGCAATTTTCCGATGCGGATACGCCCAAAATACCGTCCCTT 1114  
863 rovalSerThrGlyMetAlaSerIleValProGlyThrPheHisPro 879











```

609 ghIsthrSerGlnIleThrThrAlaProSerThrAsnHisSerThrI 626
306 TTCGGCCCTTCGACAAACATGCTT...CACATTCGGATTCTG... 343
||||| :||||| ||| :||| :|||
626 IeHisSerThrSerThrProGlnIleThrAlaValSerGlnArg 642
344 .....ACGAAGCC 351
643 GlnHisThrArgAlaProGlnIleThrGlnIleThrGlnIleThrArgSe 659
352 GGTAATCCCGTTGACGAGTTCAGCCTTTCACGATTCATTCGACGAGTA 401
659 rValSerProMetThrAsp.....Thylrst 668
402 CGAACAACATCCCGCCGACGGCTATGACG.....GGCCACAGAGCGGCG 445
668 hValThrThrProGlySerSerPheThrAlaSerGlyHisSerProSer 684
446 GCTATCCCGCTCCCAAGCGCGAGGATATATACAGCTACGACATATAA 495
685 GluIleValProGlnAspAla..... 691
496 GCGCTGCCCAAAATATCCGCTCAGCTGACGCAACCGCACGACCGG 545
692 .....ProThrIleSerAlaIleThrThrPheAlaProAlaProThrG 706
546 ACAAGCGCTTGGCAGCCGTTCCCAATGCGCGGCTATGCTGACGCAAG 595
706 lYasnGlyHisThrThrGlnIleAlaProThrThrAlaIleu..... 718
596 GAGTAGGCGAGCATTCAAAGCGCGACCGGATACGCGCGAGCTGGAC 645
719 .....GlnAlaIleAlaProSerSerHI 725
646 AGATCGGGCAATGCCCGCGAAGCCTTCACGCGACGATATCTGTCAA 695
725 sasprAlaThrIleuGlyProSerGlyGlyThrSerIeu.....SerI 739
696 AAACATCATCGCGCGCGAGCAAAATGTGCGCGCAGCGAGCTGCG 745
739 ysthGly..... 741
746 AGGATATAGCGAAGGCTCAAAACATGTGTCATGACGCGTTGGGTG 795
742 .....AlaIleuThrIleuAlaAsnSerAlaValSerThrPr 753
796 CTTTCACCGCAAAACAGATGCGCGCATCAACGATTGGCAGATATGCG 845
753 oGlyGlyProGlnIleuThrThrSerAla..... 763
846 GCAACTCAAGACTATGCGCGACGACGATCCGCGATT..... 883
764 ..SerAlaSerThrSerProAspThrAlaAlaMetThrHisThrHis 779
884 .....GGCGATCCCAAAACCCCATGCCGCA 909
780 GlnAlaIleuSerThrGlnAlaIleSerGlyIleThrGlnIleThrSerGlnProAl 796
910 CAAAGCA.....TAGAAGCGCT 926
796 aserSerGlySerArgThrThrSerAlaGlyThrAlaThrProSerSer 813
927 CAGCAATATCTTTATGCGACGATCCCATCAAGAGGATTTGAGCTGTCC 976
813 erGlyAlaSerGlyThrThrProSerGlySerGlyIleSerThrSer 829
977 GGGAAATATAGGCTTGGCGGCAATCA.....CGGACATCTCT 1014
830 GlnIleuThrThrArgPheSerSerAsnProSerArgAspSerHisThrTh 846
1015 GTCAAGCGGTCGAGATGGCGGAGATGCGATTCGCCAAAGAGAAATCCGC 1064
846 rGlnSerThrThrGlnIleuSerAlaSerHisGlyAlaIleAla 863

```

```

1065 CGTCAGGACAAATTTTGCCGATCGGCAATPACGCCAAATACCCGCTT 1114
|| :||| :||| :||| :|||
863 rovalSerThrGlyMetAlaSerSerIleValProGlyThrPheHisPro 879
1115 ACCATTCGCCAAATATCCGTT.....CAAACTTGACGACGCT 1152
||||| :||| :||| :||| :|||
880 ThrIleuSerGlnAlaSerThrAlaGlyArgProThrGlyGlnSerSerPr 896
1153 TACGGCAAAAGAAACATACCTCTCAGCGTGCAGCGCGCGCAAGCG... 1199
||||| :||||| :||| :||| :|||
896 oThrSerProSerAlaSerProGlnIleuThrAlaIleSerArgMet 912
1200 .....CAAAATGTCAACTGGCACA..... 1220
913 AlaGlnThrGlnIleArgThrGlyThrSerArgGlySerAspThrIleSerIe 929
1221 .....CCAGCGCCACCC 1232
929 uAlaSerGlnAlaThrAspThrPheSerThrValProThrProPro 945

seq_name: sp_human:Q9NY09
seq_documentation_block:
ID Q9NY09 PRELIMINARY; PRT; 2169 AA.
AC Q9NY09;
DC 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MUC1N 4.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20381033; PubMed=10920259;
RA Choudhury A., Moniaux N., Wimpenny J.P., Hollingsworth M.A.,
RA Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin cDNA and its variants in pancreatic carcinoma.";
RL J. Biochem. 128:233-243(2000).
DR EMBL; AJ276359; CAB81773.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR003886; Nidogen_ext.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00094; vwd; 2.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00539; NIDO; 1.
DR SMART; SM00216; WMD; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
SQ SEQUENCE 2169 AA; 231470 MW; 5CBA1B16AF73E469 CRC64;

```

```

alignment_scores:
  quality: 155.50      length: 517
  ratio: 0.713        gaps: 21
  percent similarity: 42.166  percent identity: 22.631

```

```

alignment_block:
us-09-303-518d-463 x Q9NY09 ..

```

```

Align seg 1/1 to: Q9NY09 from: 1 to: 2169

```

```

9 TTCGCAAAATATCCCTTATCTGTCATCT..... 41
:|||||:|||||:|||||:|||||:|||||
493 TrpProSerSerPheSerSerIleGlyHisThrThrTrpSerGlnThrGln 509
42 ...GGCAGTGTGCTCGGATGATGACACAGCGCTCAGATTGGCAAC 87
|||||:|||||:|||||:|||||:|||||
509 uIeuProSerThrSerThrGlyAlaIleAlaThrArgIleuValThrGlyAsn 526

```

```

88 GATCCCTTTATCCGGAGGTTCTCAGCCGTC..... 118
   |||
526 roserThrAlaIaIaGlyThrIleProArgValProSerLysValSer 542
119 AGCATTTCCAAACCCGAGGGAATACCACCTATTCGGCAGCAGGGGGGAG 168
   |||
543 AlalIeGlyIeProGlyIeProThrIleProSerHisSerThrIle 559
169 CTGCGCAGGCGCAGC.....GCCATATCGATTCG..... 199
   |||
559 rIeupProLysThrThrGlyAlaGlyIaIaGlyIeProThrIleGlyIe 576
200 .....GAAACATAC 208
576 hrGlyThrIleGlyIaIaIeLeuSerSerProSerLysValIle 592
209 AAGGCCATCGATTGGCGCACCTGATGATTCACACAGCGCGCTTG...AA 255
   |||
593 GImeIleLysThrAlaThrSerProSerSerSerProMetLeuAspArg 609
256 GGAATAATCGGTACATTTGTCGCTTTCCGATCAGCGGCACAAATTCGA 305
   |||
609 gHISThrSerGlnGlnIleThrThrAlaProSerThrAsnHisSerThrI 626
306 TTGCGCCCTCGACACCATGCTT...CACATTCGATTCGT..... 343
626 IeHisSerThrSerThrSerProGlnIleSerProAlaValSerGlnArg 642
344 .....ACGAGGCC 351
643 GlnHisThrArgAlaProGlnThrIleGlnIleSerGlnThrThrArg 659
352 GGTAGTCCCTGACGAGATTCACCTTTACCGATCATTTGGACGAGATA 401
   |||
659 rValSerProMetThrAsp.....ThrLysT 668
402 CGAAGACCATCCGCGCAGCGGTATGACG.....GCCACAGCGCGGAG 445
   |||
668 hrValThrThrProGlySerSerThrAlaSerGlnHisSerProSer 684
446 GCTATCCCGCTCCAAAGCGCGAGGATATATACGTAACGATGATAAA 495
   |||
685 GluIleValProGlnAspAla..... 691
496 GCGCTTGGCCAAATATCCGCTCAACCTGACGACACCGCAGACCGG 545
   |||
692 .....ProThrIleSerAlaIaIaThrThrPheAlaProAlaProThrG 706
546 ACAAGGCTTGGCCGACCGTTCCACAATTCGCGCGCTATGCTGACGCAAG 595
   |||
706 LysAngLysHisThrThrGlnAlaThrProThrAlaLeu..... 718
596 GAGTAGCGCGAGGATTCAAACGCCGCCACCGGATACAGCCCGGAGCTGGAG 645
   |||
719 .....GlnAlaIaIaProSerSerHis 725
646 AGATCGGCGCATGCGCGCAGAGCTTCACGCGGCTGACATATTCGCA 695
   |||
725 SASpAlaThrLeuGlyProSerGlyIleThrSerLeu.....SerL 739
696 AAACATCATCGCGCGCGCAGAGAAATGTGCGCGCAGGCGATCCGCTGC 745
   |||
739 ystIleGly..... 741
746 AGGCTATTAAGGAGGCTCAACACTTGTCTGTACGACGCGCTTGGGCTG 795
   |||
742 .....AlaLeuThrLeuAlaAsnSerValIaIaSerThrPr 753
796 CTTTCCACCGGAAACAGATGCGCGCATCAGCATTTGGCAGATATGCG 845
   |||
753 cGlyGlyProGlnIleGlnThrThrSerAla..... 763

```

```

846 GCAACTCAAAAGCTATGCGCGCAGCAGCATCCGAGT..... 883
   |||
764 ..SerAlaSerThrSerProAspThrAlaAlaIaMetThrHis 779
884 .....GGCGACTCCAAACCCCAATGCCGCA 909
780 GlnAlaIeSerThrGlnAlaIaSerGlyIleThrGlnThrSerGluProAl 796
910 CAAGGCA.....TAGAGCGCT 926
796 aserSerLysThrIleThrThrSerAlaGlyThrAlaThrProSerHis 813
927 CAGCAATATCTTTATGCGACCATCCCATCAAGAGGATTCGAGCTGCC 976
   |||
813 erGlyAlaSerGlyThrThrProSerGlyIleSerThrSer 829
977 GGGGAAATACGCTGGCGCGCATCA.....CGGCACATCCT 1014
   |||
830 GlyIleThrThrArgPheSerSerAsnProSerArgAspSerHisThr 846
1015 GTCAGCGCGTCCGAGATGGCGCGCATGCGCATTCGCGAAAGGAAATCCGC 1064
   |||
846 rGlnSerThrThrGlnLeuLeuSerAlaSerAlaSerHisGlyAlaIleP 863
1065 CGTCAGCGCATTTTGGCGATGCGGCATACGCCAAATACCGCTCCCTT 1114
   |||
863 rovalSerThrGlyMetAlaSerIleValProGlyThrPheHisPro 879
1115 ACCATTCGCCAAATATCCGTT.....CAAACCTGAGACCATCCT 1152
   |||
880 ThrLeuSerGlnAlaSerThrAlaGlyArgProThrGlyIleSerPr 896
1153 TACGCGCAAGAAACATCATCTCTCAACCGTCCGCGCGCTCAACAG... 1199
   |||
896 cHisSerProSerAlaSerProGlnGlu..ThrAlaAlaIaIaSerArgMet 912
1200 .....CAAAATGTCAACTGGCAGA..... 1220
913 AlagInThrGlnArgThrGlyThrSerArgIleSerAspThrIleSerIe 929
1221 .....CCACGCGCACCC 1232
929 uAlaSerGlnAlaThrAspThrPheSerThrValProProThrProPro 945

```

```

seq_name: sp_human:014881
seq_documentation block:
ID 014881 PRELIMINARY; PRT; 622 AA.
AC 014881;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MUC1N (FRAGMENT).
GN MUC5B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TRACHEOBRONCHIAL MUCOSA;
RA Desseyn J.L., Guyonnet-Duperrat V., Porchet N., Aubert J.P., Laine A.;
RT "Human mucin gene MUC5B: the 10.7 kb large central exon encodes
RT various alternate subdomains resulting in a super repeat.";
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: X74955; CAA52910.1; -.
FT NON_TER 1
FT NON_TER 1
FT SEQUENCE 622 AA; 61786 MW; 4FC85A52F50D57E5 CRC64;
SQ

```

alignment\_scores: 154.00

Length: 442

Ratio: 0.740 Gaps: 20  
Percent Similarity: 47.059 Percent Identity: 24.661  
Alignment block:  
US-09-303-518D-463 x Q14881 ..

Align seg 1/1 to: Q14881 from: 1 to: 622

```

68  ACCGCTCATGATTTGGCAACAGATCCCTTTATCCGCGAGGTTTCGACCGT 117
    |||||
187 ThrProGluThrThrHisThr.....SerThrVal 196
118 CAGCATTTCCGACCCGAGCGGAAATACCACCTATTCGCGAGCGGGA 167
    |||||
196 IleuThrThrHisAlaThr.....ThrThrArgAlaThrSerSerThrS 211
168 GCTTGGCCAGGCGCAAGGCCATATCGATGGGAAACATCAAAAGCCATC 217
    |||||
211 erThrProSerSerThrProGlyThrThrTrpIleuThrGluLeuThr 227
218 AGTTGGGCACTGATGATTCACAGCGCGCG..... 250
    |||||
228 ThrAlaAlaThrThrThrAlaGlyThrGlyProThrAlaThrProSer 244
251 .....TTGAAGAAATATCGCTACATTCGCG 278
244 rThrProGlyThrThrTrpIleuThrGluLeuThrThrAlaAlaThr 261
279 CTTTTCGATCAGGGG.....ACAAATTCATTCGCTTCGACA 319
    |||||
261 hThrAlaSerThrGlySerThrAlaThrLeuSerSerThrProGlyThr 277
320 ACCATGCGCTGACATTCGATTCGACGAGCGGATGCTCCGTTGACGGA 369
    |||||
278 ThrTrp.....IleuThrGluProSerThrThrAlaThrVal 290
370 TTCAGCCTTACCGCATTCATTCGGAGCGATACGACCATCCGCGCA 419
    |||||
290 lThrValProThrGlySerThrAlaThrAlaSerSerThr...GlnAla 366
420 CGGCTATGAGCGGCGACAGGGCGGCTATCCGCTCCA..... 460
    |||||
306 hThrAlaGlyThrProHisValSerThrThrAlaThrThrProThrVal 322
461 .....AAGCGCGAGGATATATACAGTACGACATAAAGCGCTGCC 504
    |||||
323 SerSerLysAlaThrProSerSerSerProGlyThrAlaThrAlaLeu 339
505 CAAATA...TCGCGCTCAACCTGACCGACA.....ACGCGAG 539
    |||||
339 oAlaLeuArgSerThrAlaThrThrProThrAlaThrSerPheThrAla 356
540 CACCGGACAGCGCTTGGCGACGCT.....TCGACA 571
    |||||
356 leProSerSerSerLeuGlyThrThrTrpThrArgLeuSerGlnThrThr 372
572 ATGCCGCGCTATGCTAGCGCAAGAGTAGGCGAGTCAACGCGCC 621
    |||||
373 ThrProThr.....AlaThrMetSerThrAlaThr 382
622 ACCGGATACAGCCCGAGCTGGACACATGGGCAATGCCCGCAACCTT 671
    |||||
382 rProSerSerThrProGluThrVal.....HisThrS 393
672 CAAGGCACTGAGATATCGTCAAAACATCAACGCGCGGAGGAGAA 721
    |||||
393 erThrValLeuThrAlaThrAlaThrThrThrGlyAlaThrGlySerVal 409
722 TTTCGCGAGCGGATGCCGTGACAGGATTAAGCGAGCTCAAAACAT 771
    |||||
410 AlaThrProSerSerThrProGlyThrAlaThrHisThrThrValPro 426
772 GCTGTATGAGCGCTTGGGTCTGCTTCCACCGAAACAGATGCGCG 821

```

```

426 rThrThrThrThrGlyPheThrAlaThrProSerSerSerProGlyThrA 443
    |||||
822 CATCAACGATTTGGCAGATATGGCGCACTCAAGACTATGCG..... 865
    |||||
443 lAluThrPro...ProValTrpIleSerThrThrThrProThrThr 458
866 CAGCACCATCCGCGATTTGGCGAGTCCAAACCCCATCCGCGACAGCG 915
    |||||
459 ThrThrProThrThrSerGlySerThrValThr..... 469
916 ATAGAAGCCGTCAGCAATATCTTTATGCGAGCACCATCCATCAAGGAT 965
    |||||
470 .....ProSerSerThr.....ProGlyThrThrHisThrA 480
966 TGGAGCTGTCCGGGGAATACGCGCTGGCGGCGATCAGCGACATCCG 1015
    |||||
480 lArgValLeuThrThrThrThrThrThrValAlaThrGlySerMetAla 496
1016 TCAAGCGGTGCGAGATGGCGCGATCGCATTCGCAAGAGGAATCCGC 1065
    |||||
497 ThrPro.....SerSerSerThrGlnThrSerGlyThrProPro 509
1066 CTCAGCGACATTTGGCGATGCGGCAATACG..... 1096
    |||||
509 oSerLeuThrThrThrAlaThrThrThrIleThrAlaThrGlySerThrThrA 526
1097 ..CAAATACCGCGCTCCCTTACCATTCGCGAAATATCGGTTCAACTGG 1144
    |||||
526 snProSerSerThrProGlyThrThrProIleProValLeuThrSer 542
1145 AGCAGCGTTACG.....GCAAGAAACATCACTCTCTCA 1179
    |||||
543 ThrAlaThrThrProAlaAlaThrSerSerLysAlaThrSerSerSerSe 559
1180 ACCGTGCGCGCGCTCAACGCGCAAAATGCAAAATGCGAGACGACGCA 1229
    |||||
559 rPro.....ArgThrAlaThrThrLeuProValLeuThrSerThrAla 574
1230 CCCGAAGACAGCGCTTACCGTTTGGACG 1255
    |||||
574 hLysSerThrAlaThrSerPheThr 582
seq_name: sp_fungi:Q92223
seq_documentation_block:
ID Q92223 PRELIMINARY; PRT; 961 AA.
AC Q92223;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CHITINASE.
GN CHIA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
PN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC89;
RX MEDLINE=98162139; PubMed=9501518;
RA Takaya N., Yamazaki D., Horiuchi H., Ohta A., Takagi M.:
RT Cloning and characterization of a chitinase-encoding gene (chia) from
RT Aspergillus nidulans, disruption of which decreases germination
RT frequency and hyphal growth.
RL Biosci. Biotechnol. Biochem. 62:60-65(1998).
CC -1- SIMILARITY: BELOONS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
DR EMBL: D87895; BAA36223.1.
DR HSSP: P23472; 2HW.
DR InterPro: IPR001579; Chitinase_2.
DR Pfam: PF00192; chitinase_2; 2.
DR PROSITE: PS01095; CHITINASE_18; 1.

```

KW Glycosidase; Hydrolase.  
SO SEQUENCE 961 AA; 96839 MW; 66263B3325C533E2 CRC64;

alignment\_scores:  
Quality: 152.00 Length: 436  
Ratio: 0.724 Gaps: 18  
Percent Similarity: 48.165 Percent Identity: 22.936

alignment\_block:  
US-09-303-518D-463 x Q92223 ..

Align seg 1/1 to: Q92223 from: 1 to: 961

```

97 ATCCGCGCAGTT.....CTCGACGCTGACATTTGACCCGCGG 137
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
332 ILEARGIULEUTYRASPLEUASPROADNHIS.PROPROPTHNT 348
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 GAATAACACCTATTGCGACGAGGGGGAGCTTGCCNAGCGCAAGGCC 187
|||:|||||:|||||:|||||:|||||:|||||:|||||:
348 hrserProthProthProth.....ProserThrThrThr 360
|||||:|||||:|||||:|||||:|||||:|||||:
188 ATTCGATGGGAACATCAAGCATCAGTTGGGCCACCTGATGANT 237
|||:|||||:|||||:|||||:|||||:|||||:|||||:
361 ThrSerThrThrSerThr.....ThrSerThrThrThr 366
|||||:|||||:|||||:|||||:|||||:|||||:
238 CAACAGCGCGCTTGAGGAATATCGCTACATTTGCCGTTTCCGA 287
|||||:|||||:|||||:|||||:|||||:|||||:
367 .....ThrSerThrThrSerThrThrThrThrThrThr 375
|||||:|||||:|||||:|||||:|||||:|||||:
288 TCACGGGACAAATTCATTCGCTTGCACCAACATGCTCATCATCG 337
|||||:|||||:|||||:|||||:|||||:|||||:
375 hrThrSerThrThrSerThrThrThrThrThrThrThr 386
|||||:|||||:|||||:|||||:|||||:|||||:
338 ATTCTGACGAGCGGAGTCCGCTGACGATTCAGCTTACGCGCATC 387
|||||:|||||:|||||:|||||:|||||:|||||:
387 .....ThrThrProthThrThrThrThrThrThrThrThr 398
|||||:|||||:|||||:|||||:|||||:|||||:
388 CATTGGACGAGTACG...AACACATCCCGCGACGCTATGACGGCC 434
|||:|||||:|||||:|||||:|||||:|||||:|||||:
398 rThrThrThrThrProthProthProthProthThrThrThrThr 415
|||||:|||||:|||||:|||||:|||||:|||||:
435 ACAGCGGGGGCTATCCCGCTCCCAAGCGCGAGGATATATACGCT 484
|||||:|||||:|||||:|||||:|||||:|||||:
415 hrThrGlutThrValThrProthProthProthProth 426
|||||:|||||:|||||:|||||:|||||:|||||:
485 ACGACATAAAGCGCTTGCCCAAAATATCCGCTCAACCTGACGCAAC 534
|||||:|||||:|||||:|||||:|||||:|||||:
427 .....ProserGlutSerThrThrSerGlutThrSe 437
|||||:|||||:|||||:|||||:|||||:|||||:
535 CGCAGACCGGACACGCGCTTGCCGACCGTTTCCACATGCGCGCTAT 584
|||||:|||||:|||||:|||||:|||||:|||||:
437 rSerLeuProthThrThrSerThrThrProthValSerGlutThr 452
|||||:|||||:|||||:|||||:|||||:|||||:
585 GCTGACCGCAGAGTAGCGAGGATTCAAAGCGCCACCC..... 625
|||||:|||||:|||||:|||||:|||||:|||||:
453 .....GlutThrLysThrProthThrSerThrAlaProthLeuSer 467
|||||:|||||:|||||:|||||:|||||:|||||:
626 .....GATACAGC 633
|||||:|||||:|||||:|||||:|||||:|||||:
468 SerSerProthValGlyGlySerSerThrAlaSerSerThrThrThr 484
|||||:|||||:|||||:|||||:|||||:|||||:
634 CCGCAGCTGACAGATCGGCAATGCCCGCGAAG.....CCTTCAACGG 677
|||||:|||||:|||||:|||||:|||||:|||||:
484 rProserGlutThrProserAlaSerSerThrAlaValSerGlutThr 501
|||||:|||||:|||||:|||||:|||||:|||||:
678 CACGCGCATATGCTCAAAACATCAATGCGCGGCGAGGAGGAATG... 724
|||||:|||||:|||||:|||||:|||||:|||||:
501 erThrHisLysSerThrThrSerThrGlyProthLysThrThrThr 517
|||||:|||||:|||||:|||||:|||||:|||||:
725 .....TCGCGCAGCGGATCGCTGACGAGGTAAAGCGAAGGCTCAACAT 771
|||||:|||||:|||||:|||||:|||||:|||||:
518 GlySerSerThrSerValProthAlaThrSerSerSerValProserSer 544
|||||:|||||:|||||:|||||:|||||:|||||:

```

```

772 GCTGTCAATGACGCGCTTGCTGCTTT.....CCACCGA 806
|||||:|||||:|||||:|||||:|||||:|||||:
534 alSerProthSerThrProthValLysSerGlutThrProthProthProth 551
|||||:|||||:|||||:|||||:|||||:|||||:
807 AACACAGATGCGCGCATCAACGATTTGCGACATATGCGCACTCAAG 856
|||||:|||||:|||||:|||||:|||||:|||||:
551 alThrSerSerSerSerThrThrValSerSerThrThrThrThr 567
|||||:|||||:|||||:|||||:|||||:|||||:
857 ACATGCGCAGCAGCAGCATCCGCGATTTGGCAG..... 889
|||||:|||||:|||||:|||||:|||||:|||||:
568 AspysSerGlutSerSerThrAlaLysLysThrHisSerSerSeril 584
|||||:|||||:|||||:|||||:|||||:|||||:
890 .TCCAAACCCCAATGCGCGCAGCAGCATGAAAGCGTCAATATCTT 938
|||||:|||||:|||||:|||||:|||||:|||||:
584 eSerGlutThrPro.....SerAlaSerThrProthAlaSerPro 598
|||||:|||||:|||||:|||||:|||||:|||||:
939 TATGCGAGCATCCCATCAAGGATTTGAGCGT..... 973
|||||:|||||:|||||:|||||:|||||:|||||:
598 erThrSerProthThrThrLysThrLeuThrValPheProthProth 614
|||||:|||||:|||||:|||||:|||||:|||||:
974 TCCGCGGAAATACGCGCTTGCGCGCATCAGCGCACATCTGCAAGCG 1023
|||||:|||||:|||||:|||||:|||||:|||||:
615 SerSerValSerThrThrThrThrSerAlaSerThrLeuSerSerVa 631
|||||:|||||:|||||:|||||:|||||:|||||:
1024 TCGCAGATGGCGCGATCGCATTTCCGCAAGGAAATCCCGCGCAGCA 1073
|||||:|||||:|||||:|||||:|||||:|||||:
631 lProthThrSerLysLysThrThrThrThrThrThrThrThrThr 648
|||||:|||||:|||||:|||||:|||||:|||||:
1074 CAATTTGCCGATCGGCATACGCCAATACCCGCTTACCATTTCC 1123
|||||:|||||:|||||:|||||:|||||:|||||:
648 erAlaSerGlutThrProserAlaSerThrSerLysProth... 663
|||||:|||||:|||||:|||||:|||||:|||||:
1124 GAATATCCGTTCAACTGAGC.....AGCTTACGCGCAAGAAAC 1167
|||||:|||||:|||||:|||||:|||||:|||||:
664 ThrsSerAlaSerSerThrThrThrThrThrThrThrThrThr 680
|||||:|||||:|||||:|||||:|||||:|||||:
1168 ATCACTCTCTCAACCGTGGCGCGCT.....CAACGCAAAA 1205
|||||:|||||:|||||:|||||:|||||:|||||:
680 rAlaProthValProserSerSerAlaGlySerAlaSerThrThrThr 697
|||||:|||||:|||||:|||||:|||||:|||||:
1206 TGTCACTGCGACAGCAACCGCCACCGCAAGCGGTACCGTTGACG 1255
|||||:|||||:|||||:|||||:|||||:|||||:
697 hrSerThrThrThrThrThrThrThrThrThrThrThrThrThr 713
|||||:|||||:|||||:|||||:|||||:|||||:
1256 GTAAG 1261
|||||:|||||:|||||:|||||:|||||:|||||:
714 Allys 715
|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: sp_rodent:070495
seq_documentation_block:
ID 070495 PRELIMINARY; PRT; 897 AA.
AC 070495;
DT 01-AUG-1998 (TEMBLrel. 07, Created)
DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE PLENTY-OF-PROLINES-101.
GN SRMT1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN.
RA Vayssiere B.M., Camonis J.H.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF062655; Acl17422.1;
DR MGD; MGI:1858303; Strml.
DR InterPro; IPR002483; PWT.
DR Pfam; PF01480; PWT.1.
DR SMART; SMO0311; PWT.1.

```





RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL, AB016091, BAB3717.1, -  
FT MONTER 1  
SQ SEQUENCE 1275 AA; 136869 MW; 45C2BF85E98A6F6 CRC64;

## alignment\_scores:

Quality: 150.00 Length: 598  
Ratio: 0.558 Gaps: 28  
Percent Similarity: 44.983 Percent Identity: 22.742

## alignment\_block:

US-09-303-518D-463 x Q9UQ36 ..

Align seg 1/1 to: Q9UQ36 from: 1 to: 1275

```
39 ACTGCGACGTGTGCT...GCCGATGCATGCACAGCCCTCAGATTGGCAA 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
102 SerGlySerSerProGluValAspSerLysSerArgLysSer...ProAr 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
86 ACATCCCTTATCCGCGACGTTCTGACCCCTCAGCATTCGACCCGAC 135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117 gArgSer...ArgSerLysSerSerProGluValLysAspLysProA 132
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
136 GGGAAATACCACTATTCGCGACAGGCGGAGCTTCGCGACGCAACG 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
132 rGalaIaPro...ArgAlaGlnSerGlySerSpsSerSerProGlu 146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
186 CCATATCGATTGGGAAACATACAAACCATCATCGTGGCCACCTGATGA 235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
147 ProLysAlaProAlaProAlaGlnProAlaArgSerArgSerGlySe 163
236 TTCACACGGGGCGCTTGAAGAAATATGGCTACATTCGCGCTTTC 285
163 rSerSerLysGlyArg...GlyProSerProG 173
286 GATCAGCGGCGACAAATTCATTCGCTTCGACACACATGCTCACATTC 335
173 LysLysSerSerThrLysSerSerProGluHisProPro... 186
336 CGATTTCAGCAAGCCGCTAGTCCCGTTGACGATTACGCTTACCGCA 385
187 ..LysSerArgThrAlaArgArgGlySerArgSerSerProG 200
386 TTCATTGGGCGGATACGACACATCCGCGCGAGCTGTCAGGGGCA 435
200 uProLysThrLysSerArgThrProProAlaArgArgSerSerArgS 217
436 CAGGCGCG...CGGCTATCCGCTCCCAAGG... 464
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 erProGluLeuThrArgLysAlaArgLysSerArgSerArgSerAla 233
465 ..CGGAGGATATATACAGTACGACATCAATAAGGCG 499
234 SerSerSerProGluThrArgSerArgThrProProAlaHisArgSe 250
500 TTGCGCAAAATATCCGCTCAGCTACGACAA... 533
250 rProSerValSerSerProGluProAlaGlnLysSerArgSerSerArg 267
534 ..CCGACGACCGGACACAGCGCTTGCCGA... 560
267 rArgArgSerSerAlaSerSerProArgThrLysThrSerArgArgGly 283
561 ..CGTTTCACAAATGCGGCGCTATGTCGACGACAGGAGTAGG... 602
284 ArgSerProSerProLysProArgGlyLeuGlnArgSerArgSerArgSe 300
603 ..CGACGGATTCAACGCGCCACCGG...ATACAGCCCGGAGCTGACAGA 648
300 rArgArgGlnLysThrArgThrArgArgArgArgArgSerGly... 315
649 TCGGCAATGCGCGCAAGCTTCAACGCGACATCAATATCGTCAAAA 698
```

```
316 .....SerSerGlnSerThrSerArgArgArgGlnArg 326
699 CATCATCGGCGCGGACAGAGAAATTCGCGGACGCGGATCGCGTAGG 748
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
327 SerArgSerArgSerArgValThr...ArgArgArgArgGlyGlySerG 342
749 GTATATACCGAAGGCTCAACATTCGCTGTCATGACGCGCTTGCTGCTT 798
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342 YTr...HisSerArg... 346
799 TCACACGAAACAAAGATGCGCGCATCAGCATTTGGCAATATGGCGCA 848
347 .....SerProAlaArgGlnLysSerSerArgThrSerSer 358
849 ACTCAAGACATATGCGGACGACCATCCGATTTGGCGATCCCAAAAC 898
359 ArgArgArgArgArgGlyArgSerArgThrPro...Pr 369
899 CCAATGCGCGACAAAGGACATAGACCGCTCAGCAATATCTTATGGCAGC 948
369 oThrSerArgLysArgSerArgSerArg...ThrS 380
949 ATCCC...CATCAAGGATTTGACCTGCTCCGCGGGAATA 986
380 erProAlaProThrLysArgSerArgSerArgAlaSerProAlaThrHis 396
987 CGGCTTGGGCGGACATCAAGCGCATCTCTGTCACAGGCTGCGAGTGGCG 1036
397 ArgArgSerArg...SerArgThrProLeuLysSerArgArgArg 410
1037 CGATGCGATTCGCGAAAGGAAATCCGCGCTCAGCGACATTTTGGCGAT 1086
410 gSerArgSerArgThrSerProValSerArgArgArgSerArgSerArgT 427
1087 GCGGATACGCGCAATACCCGCTCCCTTACCATTCGCGAATATCCGTC 1136
427 hrSerVal... 429
1137 AAACCTTGAGCAGCGTTAGCGCAAAAGAAACATCACCCTCAACCGTGC 1186
430 .....ThrArgArgArgSerArgSerArgAlaSerProVa 441
1187 CGCGCTCAAGCGCAAAATGTCAAACTGCGACAGCAACGCCACCGCAG 1236
441 lSerArgArgSerArgSerArg...ThrProProValT 454
1237 ACAGCGTACCGTTGACGCGTAAAGGCTTCCGAATTTGAGAAGCACT 1286
454 hrArgArgArgSer...ArgSerArgThrProThrThrArgArgArgSe 469
1287 GAAATATGATACGAAGCTCGATATTCGAATTTATCGGGGCGGTATAC 1336
469 rArgSerArgThrProProValThrArgArgArgSerArgSerArgThrP 486
1337 CTAAGGCTAAGCCTGTGTTGATGCGAAACGAGATGGGAGGTGATAGG 1386
486 ro...ProValThrArgArgArgSerArg...Ser 495
1387 AAGCTTAATTAATTGACAACTGTCAGCAGGTGCGAATAATGTTTCAGA 1436
496 ArgThrSerProIleThrArgArgArgSerArgSerArgThrSerProVa 512
1437 AACGAGAAAGAGACTCAGAGTAGTCAGTTAAAGCCATGCGCAGAG 1486
512 lThrArgArgArgSerArgSerArgThrSerProValThrArgArgArg 528
1487 AATGGAAATAAACAAGGTTAGATTATTCATTATTAAGTGGTGTAT 1536
529 .....Ser 529
1537 ATCAATTAAGAAAGACAGTAACAGAGGCGATAGTCAATCCCGTGTA 1586
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

530 ArgSerArgThrSerProValThrArgArgSerArgSerArgThrPr 546  
 1587 TGTACGGGTGATACAAACCTCGCA.....CCTGATA 1621  
 546 oProAlaIleArgArgSerArgSerArgThrProIleuProArgL 563  
 1622 AACATGGGGTTTATCAACGACAGTGAATTTAAAAAGCTGATGAAGT 1671  
 563 ySarSerArgSerArgSerProIleuAlaIleArgArgSerArgSer 579  
 1672 TGGAGGTGAACGAAAAAGGTGGGAAGTGAACCAAG 1713  
 580 ArgThrProArgThrAlaArgGlyLysArgSerLeuThrArg 593

seq\_name: sp\_human:015038

seq\_documentation\_block:  
 ID 015038 PRELIMINARY; PRT: 1783 AA.  
 AC 015038;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)  
 DE KIAA0324 PROTEIN (FRAGMENT).  
 GN KIAA0324.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE-97349984; PubMed-9205841;  
 RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT Prediction of the coding sequences of unidentified human genes. VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 4:141-150(1997).  
 DR EMBL: AB002322; BAA20782.2; -;  
 FT NON\_TER  
 FT SEQUENCE 1783 AA: 190940 MW: 6603026FD4179AB CRC64;

alignment\_scores:

Quality: 150.00 Length: 598  
 Ratio: 0.558 Gaps: 28  
 Percent Similarity: 44.983 Percent Identity: 22.742

alignment\_block:

US-09-303-518D-463 x 015038

Align seg 1/1 to: 015038 from: 1 to: 1783

39 ACTGGCAGTGTGCT...GCCGATGATGCACAGCCTCATGATTGGCA 85  
 610 SerGlySerSerProGluValAspSerGlySerArgLeuSer...ProAl 625  
 86 ACATTCCTTATTCGCGAGGTCTCGACCGTCAGCATTTGCAACCCGAC 135  
 625 gArgSer.....ArgSerGlySerSerProGluValLysAspLysPro 640  
 136 GGGAAATACCACTATTCGCGACAGAGGGGAGCTTCGCCNAGCGCAACG 185  
 640 rGAlaAlaPro.....ArgAlaGlnSerGlySerAspSerSerProGlu 654  
 186 CCATATCGATTGGGAACATACAAAGCCATGATTGGGCACACCTATGA 235  
 655 ProGlyAlaProAlaProAlaLeuProAlaLeuProAlaArgSerArgSerGly 671  
 236 TTCACAGCGCGCGCTTGAAGAATATCGGCTACATTGTCGCTTTTCC 285  
 671 rSerSerGlyGlyArg.....GlyProSerProG 681

286 GATCAGCGGCAAAATTCATTCGCCCCCTTCGACACCATGCCATTC 335  
 681 IuGlySerSerSerThrGluSerSerProGluHisPro..... 694  
 336 CGATTTCAGCAAGCCGGTAGTCCCTTGACGGATTCAGCCTTACCGCA 385  
 695 .....LysSerArgThrAlaArgArgGlySerArgSerProG 708  
 386 TTCATTGGAGCATACAGACACCATCCGCCACGCGCTATGACGGCCA 435  
 708 uProLysThrLysSerArgThrProProAlaArgArgSerSerArgS 725  
 436 CAGGCGCG.....CGGCTATCCCGCTCCCAAG..... 464  
 725 erProGluLeuThrArgLysAlaArgLeuSerArgSerArgSerAla 741  
 465 .....CGCGAGGATATATACACCTTCAGCATTAAGGCG 499  
 742 SerSerSerProGluThrArgSerArgThrProProAlaHisArgArg 758  
 500 TTCGCCAAATATCCGCTCAACCTGACCGCA..... 533  
 758 rProSerValSerSerProGluProAlaGluLysSerArgSerArg 775  
 534 .....CCGACACCGGCAACAGCGCTTGCGCA... 560  
 775 rArgArgSerAlaSerSerProArgThrLysThrThrSerArgArgGly 791  
 561 .....CGCTTTCACATATGCGCGCTATGCTGACGCAAGAGTAGT... 602  
 792 ArgSerProSerProLysProArgGlyLeuGlnArgSerArgSerArg 808  
 603 ..CGAGATTCAAGCGCCACCGC...ATACAGCCCGACGCTGACAGCA 648  
 808 rArgArgGluLysThrArgThrArgArgArgSpsArgSerGly... 823  
 649 TCGGGCAATCCGCGCAAGCCTTCAACGCGACTGCAGATATGTCAAAA 698  
 824 .....SerSerGlnSerThrSerArgArgArgGlnArg 834  
 699 CATCATCGCGCGCGGCAAGAAATGTGCGCGCAGCGCATGCGTCAG 748  
 835 SerArgSerArgSerArgValThr...ArgArgArgArgGlySerL 850  
 749 GTATAAGCAAGGCTCAAAATTCGTCATGCAATGACGAGCTGCTGCTT 798  
 850 yTyr.....HisSerArg..... 854  
 799 TCCACCGAAACAGATGCGCGCATCAACGATTTGGCAGATATGCGCA 848  
 855 .....SerProAlaArgGlnGluSerSerArgThrSerSer 866  
 849 ACTCAAGACTATGCCGACAGCCATCCGCGATTTGGCGAGTCAAAAC 898  
 867 ArgArgArgArgGlyArgSerArgThrPro.....Pr 877  
 899 CCAATGCCGACACAGCATAGAACCGTCAGCAATATCTTATGSCACC 948  
 877 oThrSerArgLysArgSerArgSerArg.....ThrS 888  
 949 ATCC.....CATCAAGGATTTGAGAGTGTCCGCGAATAATA 986  
 888 erProAlaProThrLysArgSerArgSerArgAlaSerProAlaThrHis 904  
 987 CGGCTTGGCGGCAATCAGCGCATCTCTGAAGCGGTGCGAGATGGCG 1036  
 905 ArgArgSerArg.....SerArgThrProLeuLieserArgArg 918  
 1037 CGATCGCATTCGCGAAAGGAAATCCCGGACGAGCAATTTTCCGAT 1086  
 918 gSerArgSerArgThrSerProValSerArgArgArgSerArgSerArgT 935  
 1087 GCGGATACGCCAAATACCCGCTCCCTTACCATTCGCAATATACGTC 1136

```

935 hSerial..... 937
1137 AACCTTGAGCAGCGTTACGCGAAGAAATCACCCTCCACCGTGC 1186
938 .....ThraArgArgSerArgSerArgAlaSerProva 949
1187 CCGCCGCAAGCGCAAAATGTCAAACGTGCGAGACCGCCGCGAAG 1236
949 lSerArgArgSerArgSerArg.....ThraProProva 962
1237 ACAGCGGTACCGTTGACGCGTAAGGGTTCCGAATTTTGAGACACGT 1286
962 hArgArgArgSer...ArgSerArgThr.ProthThraArgArgSe 977
1287 GAAATATGATACGAGCTCGATATTCAGATATTCGGGGCGGTATAC 1336
977 rArgSerArgThraProProvaThraArgArgSerArgSerArgThra 994
1337 CTAAAGCTAAAGCCTGTTGATGCGAAGCGATGGGAGGTGTATAG 1386
994 ro.....ProvaThraArgArgSerArg.....Ser 1003
1387 AACCTTAATTAATTTGACAACTCGTGACAGTGAGAAAATGTTCCAGA 1436
1004 ArgThraSerProthThraArgArgSerArgSerArgThraSerProva 1020
1437 AACGAGACAGAGATCAGACAGTACGATTAAAGCCCGCGCAACGAG 1486
1020 lThraArgArgSerArgSerArgThraSerProvaThraArgArgArg 1036
1487 AATGGGAAATTAACAGGGTTAGATTATTAATCATTTTATAGGTGAT 1536
1037 .....Ser 1037
1537 ATCAATAGAAAGCAGACAGTACAGAGAGGCGATGTAACCGGTGTA 1586
1038 ArgSerArgThraSerProvaThraArgArgSerArgSerArgThra 1054
1587 TGTACGGGTATACAAACAACTCGCGCA.....CCGANA 1621
1054 oProAlaAlaArgArgArgSerArgThraProleuLeuProAla 1071
1622 AACATGGGGTTTATCAAGCAGCAGATGGAATTAAGCCTGATGGAAGT 1671
1071 ysaArgSerArgSerArgSerProleuAlaAlaArgArgArgSerArg 1087
1672 TGGGAGGTGAAGAAAGAGTGGGAAGTATGATCCAG 1713
1088 ArgThraProArgThraAlaArgGlyLysArgSerLeuThraArg 1101
seq_name: sp_human:060382
seq_documentation_block:
ID 060382 PRELIMINARY; PRT: 1791 AA.
AC 060382;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE KIAA0324 (FRAGMENT),
GN KIAA0324,
OS Homo sapiens (Human),
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rieke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,
RA Goodwin L., Bryant J., Tesmer J., Melnick L., Longmire J., White S.,
RA Deng S., Tatum O., Campbell C., Fawcett J., Deaven L.,
RT "Sequencing of Human Chromosome 16p13.3",
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.

```

```

RN [2]
RP SEQUENCE FROM N.A.
RA Rieke D.O.;
RT "Large Scale Sequence Analysis and Annotation with the Sequence
RT Comparison Analysis (SCAN) System.";
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC004453; AAC08453.1; -.
FT NON_TER 1
FT SEQUENCE 1791 AA; 191306 MW; 3A7B5530AEB95F3E CRC64;

```

```

alignment_scores:
  Quality: 150.00 Length: 598
  Ratio: 0.558 Gaps: 28
Percent Similarity: 44.983 Percent Identity: 22.742

```

```

alignment_block:
US-09-303-518D-463 x 060382 ..

```

```

Align seg 1/1 to: 060382 from: 1 to: 1791

```

```

39 ACTGACAGTGTGCT...GCCGATCATGCACAGCGCTCAGATTGGCAA 85
611 SerGlySerSerProGluValAspSerLysSerArgLeuSer...ProAr 626
86 ACGATCCCTTTATCCGCGAGGCTTCGACCGTCAGCATTTCCGAACCGGAC 135
626 ArgSer.....ArgSerGlySerSerProGluValAspLysProA 641
136 GGGAAATACCACTATTCGCGACGAGGGGAGGCTCCNAGCCAGCAG 185
641 rGAlaAlaPro.....ArgAlaInSerGlySerAspSerSerProGlu 655
186 CCATATCGATTGGGAAACATACAAACCATCAGTTGGCGCACCTGTATGA 235
656 ProLysAlaProAlaProAlaProAlaLeuProAlaArgArgSerGlySe 672
236 TTCACAGCGCGCGGTGAAGAAATATCGCTACATTGTCCTTCC 285
672 rSerSerLysGlyArg.....GlyProSerProG 682
286 GATCAGCGGACAAATTCATTCGCCCTCGACAAACCATGCTCAGATTC 335
682 lGlyLysSerSerThraLysSerSerProGluHisProPro..... 695
336 CGATTCTGACGAGCGGCTAGTCCGTTGACGATTCAGCTTTACCGCA 385
696 .....LysSerArgThraAlaArgGlySerArgSerSerProG 709
386 TCCATTGGGACGATACGACACCATCCCGCGAGCGGTATGACGGGCA 435
709 uProLysThraLysSerArgThraProAlaArgArgArgSerArgSer 726
436 CAGGCGG.....CGGCTATCCCGCTCCCAAG..... 464
726 erProGluLeuThraArgLysAlaArgLeuSerArgArgSerArgAla 742
465 .....CGGAGGAGATATATACAGTACGACATTAAGCGG 499
743 SerSerSerProGluThraArgSerArgThraProProAlaHisArgArgSe 759
500 TTGCCCAAAATATCCGCTCAACCTGACCGCA..... 533
534 .....CCGACGACCGGACAGCGCTTGCCGA... 560
776 rGArgArgSerAlaSerSerProArgThraLysThraSerArgArgGly 792
561 .....CCGTTCCCAATGCCGCGCATGCTACAGCAAGAGAGAG.. 602
793 ArgSerProSerProLysProArgGlyLeuGlnArgSerArgSerArgSe 809

```



```

229 CTGATGATTCACAGCGCGCGTGAAGAAATATCGCTACATTGCGG 278
114 r.....ThrLeuProProHisSerSers 122
279 GTTTCGATCAGCGGACAAATTCATCCGCTTCGACAAACATG... 325
122 eTLeuProThrThrGluThrAlaThrMetThrProThrThrLeuLe 138
326 .....CCTCAC.....ATTCGATTCGACGAGAACCC 351
139 ThrThrThrProAsnThrThrSerHisSerThrProSerPheThrSe 155
352 GGTATCCCGTTCGAGGATTCAGCCTTACC..... 382
155 rThrIleYSerThrValSerThrSerThrAlaIleSerSeraIas 172
383 .....GCATCC 388
172 eProThrSerGlyThrMetValThrSerThrThrMetThrProSerSer 188
389 ATTGGAGCGATACAGACACCATCCCGCGAGCGGTATG..... 427
189 LeuSerThrAspThrProSerThrThrProThrIleThrTyProSe 205
428 .....ACGGGCGACAGCGCGCGGTATCCGCTC..... 457
205 rValGlySerThrGlyPheLeuThrThrAlaThrAspLeuThrSerThp 222
458 .....CCAAAGCGG 466
222 heThrValSerSerSerSerAlaMetSerlySerValIleProSerSer 238
467 CGAGGATATATACACTACGACATAAAGCGTGGCCCAAAATATPCGC 516
239 ProSerIleGlnAsnThrGluThrSerLeuValSerMetThrSeraI 255
517 CTCACAC.....TGACGACAAACGCGACACCGGCAACGCGCTTG. 556
255 eThrThrProSerLeuArgProThrIleThrSerThrAspSerThrLeuT 272
556 ..... 556
272 hrSerSerLeuLeuThrThrPheProSerThrTySerPheSerSerSer 288
557 .....CCGACCGTTTCACAAAT 573
289 MeSerAlaSerSerAlaGlyThrThrHisThrGluThrIleSerSerie 305
574 GCGGCGGTATGCTGACGACGAGGATGCGGATTCAAACGCGCCAC 623
305 uproAlaSerThrAsnThrIleHisThrThrIleGluSerAlaLeuAlaP 322
624 CGGATACAGCCCGAGCTGACGAGATCGGCGCAATGCCGCGGAAGCTTCA 673
322 roThrThrThrThrSerPheThrThrSerProThrMetIuProProSer 338
674 ACGGACATCGAGATATCGTCAAAAACATCATCGCGCGGACGAGAAAT 723
339 ThrThrValAlaThrThrGly..... 345
724 GTGCGCGCAGGATCGCGTGCAGGGTATTAAGGAAGCTCAAAACATTCG 773
345 ..... 345
774 TGTTCATCAGCGCTGGTCTGCTTCCACCGAAACAGATGCGCGCA 823
346 .....ThrGlyGlnThrThrPheProSerSerThr..... 355
824 TCAACGATTTGGCAGATGCGCAACTCAAGACGTATGCCGACGACGCC 873
356 AlaThrPheLeuGluThrThrThrLeuThrProThrThrAspPheSerTh 372
874 ATCGCGGATTTGGGCGATCGCAAAACCCCAATGCGCAGCAAGCATAGAAG 923

```

```

372 rglUserLeuThrThrAlaMetThrSerThrPro.....P 384
924 CGTCAGCATATGCTTTATGACGACCATCCCATCAAGAGGATTCGACG 973
384 rollerThrSerSeraIe.....ThrProThrAspThrMetThrSer 397
974 TCCGGGAAATATGAGGCTTGGCGGCATCAGCGCACATCCGTTCAGCGG 1023
398 MeTarGThrThrThrSerThrProThrAlaThrAsnThrLeuSerPro.. 413
1024 TCCGACATGGCGCGCATTCGCATTCGCAAGGAAATCCGCGTCAGCA 1073
414 .....LeuThrSerSerIleLeuSerSerThrProValProSerT 427
1074 CAATTTGCCGATGCGGATACGCCCAATATCCGCTCCCTT..... 1114
427 hrGluValThrThrSerThrThrAsnThrAsnProValSerThrLeu 443
1115 .....ACCATCC...GAATATCCGTTCAAACTTGAGACGCGTTAC 1155
444 ValThrThrLeuProIleThrIleThrArgSerThrLeuThrSerGluTh 460
1156 GCGAAGAAACATACACTCCCTCAACCGTCCCGCGCTCAAGCGCAAAA 1205
460 rAlaTyProSerSerProThrSerThrValThrGluSerThrThrGluI 477
1206 TGTCA 1210
477 leThr 478
seq_name: sp_human:Q9NYE4
seq_documentation_block:
ID Q9NYE4 PRELIMINARY; PRT; 716 AA.
AC Q9NYE4:
DT 01-OCT-2000 (TREMBLrel. 15' Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)
DE CERVIICAL MUCIN MUC5B (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ho S.B., Toribara N.W., Anway R.E., Spurr-Michaud S.J., Shekels L.L.,
RA Keulmann H.T., Hill J.A., Gipson I.K.;
RT "Expression cloning of human cervical proteins using an antibody to
RT cervical mucus."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253521; AAF64523.1; -.
FT NON_TER 1 1
FT NON_TER 716 716
SQ SEQUENCE 716 AA; 71194 MW; 6F3F20E7512289F3 CRC64;

```

alignment\_scores:

Quality:	Length:
Ratio: 148.00	436
Percent Similarity: 46.101	Gaps: 24
	Percent Identity: 23.394

alignment\_block:

US-09-303-518D-463 x Q9NYE4 ..

Align seg 1/1 to: Q9NYE4 from: 1 to: 716

```

26 TTATCTGTCCATAGTGGAGTGTGCTGCGGATGATGACAGCGCTCA 75
311 ValPheCys.....CysAsnTyGlnHisCysPro.. 320
76 GATTTGGCAAGAGATCCCTTATCGCGGAGGTTCTCGACCGTCAGCATTT 125

```

```

321 .....S 321
126 CGAACCCGACGGAAATACACCTATTCGCGACGAGGGGAGCTCCN 175
||||| .....
321 erThrAlaThrSerSerThr.....AlaMetPro 331
176 AGCGCAAGCGCATTCGATTGGGAAACATACAAAGCATCAGTTGGC 225
||||| .....
332 SerSerThrProGlyThrThrTrp..... 339
226 CACCTGATGATTCACAGCGCGCTGAGAAATATCGCTACATTGT 275
||||| .....
340 .....IleuThrGluLeuThrThrAla 349
276 CGGCTTTCCGATCAGCGGACAAATTCATTCGCCCTTGACAC... 319
||||| .....
349 hrThrThrAlaSerThrGlySerThrAlaThrProSerSerThrProGly 365
||||| .....
320 ACCATGCCCTACATTCGATTCTGACGAGCCGGTAGTCCCGTTGACGGA 369
||||| .....
366 ThrAlaProProProlValLeuThrSerProAla.....ThrTh 379
370 TTCAGCCTTACCGCATCCATTCGATGGAGGATACGACATCCCGCGA 419
||||| .....
379 rProThrAlaThrSerSerThrAlaThrSerSerSer...ProArgT 395
420 CGGCTATGACGGGCCACAGGGCGGCTATCCCGCTCCCAAGGGCGCA 469
||||| .....
395 hrAlaThrThr.....LeuProValLeuThr 403
470 GGGATATATACAGCTACGACATAAAGCGTTGCCCAATATCCGCTC 519
||||| .....
404 SerThrAlaThrLysSerThrAlaThrSerValThrProLeuProSerSe 420
||||| .....
520 AACCTGACGACACCGCAGACCGGACAAAGCGCTTCCGACCGTTCCA 569
||||| .....
420 rThrLeuGlyThrThrGlyThr.....LeuProGluGluThrT 433
570 CAATGCCGGCGTATGCTGACGAGAGAGTAGGCGAGGATTGCAACGGG 619
||||| .....
433 hrThrProVal.....AlaThrMetSerThrIle 442
620 CCACCCGATACAGCCCGGACGTGACAGATCGGGCAATCGCCGCAAGCC 669
||||| .....
443 HisProSerSerThrProGluThrThr.....HisTh 453
670 TTCACGGCAGCTGCAGATATCGTCAAAAATCATTCGGCGCGGACGAGA 719
||||| .....
453 rSerThrValLeuThrThrGluAlaThrThrThrAlaThrSerSerT 470
720 AATGTGCGGCGAGCGGATGCCG.....TGCAGGGTA.....TAA 754
||||| .....
470 hrSerThrProSerSerThrProGlyThrThrCysIleLeuThrGluLeu 486
755 GCGAAGGCTCAAAATTCGCTGCATCAGCGCTTGGGTGCTTCCACCC 804
||||| .....
487 ThrThrAlaAlaThrThrThrAlaAlaThrGlyProThrAlaMetPro 503
||||| .....
805 GAAACACAGATGCGCGCATCAACGATTTGGCAGATATGGCCCACTCA 854
||||| .....
503 rSerThr.....ProGlyThrThrTrp...IleLeuThrGluLeuT 516
855 AGACTATGCGCGGACGAGCATCGCGATTGGCGAGTCCAAACCCCAATG 904
||||| .....
516 hrThrThrAlaThrThrThrGluSerThrGlySerThrAlaThrPro... 531
905 CCGCACAGGSCATAGAACCGCTCAGCAATATCTTATGG.....CA 945
||||| .....
532 .....SerSerProGlyThrThrTrpIleLeuThrGlu 542
946 GCCATCCCATCAAAAGGATGAGCTCCGCGGAAATACGAGCTGGG 995
||||| .....
542 uProSerThrThrAlaThrValThrValProThrGlySerThrAlaThr 559

```

```

996 CGGCATACAGG.....CACATCCTCTCAAGCGGTGCG 1027
||||| .....
559 LaserSerThrGluAlaGlyThrProHisValSerThrThrAlaThr 575
1028 AGATGGCGCATCGCATTCGCGA.....AGGGAATCCGCGCTC 1068
||||| .....
576 ThrProThrValThrSerSerThrLysAlaThrProSerSerSerProGlyTh 592
1069 AGGACCAATTTTGGCG...ATCGCGATACGGCAATATACCGCTCCGCTTA 1115
||||| .....
592 rAlaThrAlaLeuProAlaLeuArgSerThrAlaThrThrProThrAlaT 609
1116 CC.....ATTCGCAAAATATCCGTT...CAAACTTGAGACAGC 1150
||||| .....
609 hrSerPheThrAlaIleProSerSerSerLeuGlyThrThrThrPhrArg 625
1151 GTTAGCGCAAGAAACATCACCCTCTCAACCGTCCGCGCGTCAACAGCG 1200
||||| .....
626 LeuSer.....GlnThrThrThrPro.....ThrAl 634
1201 AAAATGTCAACCTGACAGACCAAGCCAGCCGAGACAGCGGTACCGTT 1250
||||| .....
634 arThrMetSerThrAlaThrThrProSerSerThrProGluThrValHisThr 651
1251 TGACGGTA 1258
||||| .....
651 erThrVal 653

```

```

seq_name: sp_bacteria:Q924N7
seq_documentation_block:
ID Q924N7 PRELIMINARY; PRT; 1873 AA.
AC Q924N7;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SURFACE PROTEIN PRECURSOR.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacilli/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MHM594;
RX MEDLINE=99081742; Pubmed=9864215;
RA Shankar V., Baghdayan A.S., Huycke M.M., Lindahl G., Gilmore M.S.;
RT "Infection-derived Enterococcus faecalis strains are enriched in esp,
a gene encoding a novel surface protein."
RL Infect. Immun. 67:193-200(1999).
DR EMBL: AF034779; AAD09858.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
KW signal.
FT SIGNAL.
FT CHAIN. 50 1873 SURFACE PROTEIN
SQ SEQUENCE 1873 AA; 202084 MW; F609483DB980621 CRC64;

```

```

alignment_scores:
Quality: 148.00 Length: 695
Ratio: 0.468 Gaps: 36
Percent Similarity: 45.468 Percent Identity: 21.871

```

alignment\_block:

```

US-09-303-518D-463 x Q924N7 ..
Align seg 1/1 to: Q924N7 from: 1 to: 1873

```

```

80 TGGCAAAAGATCCCTTATCCGCGAGTTCG..... 112
||||| .....
604 TrpLysThrGluProAspThrSerAsnValThrAspSerThrLysGlyI 620

```

```
113 ... ACCGTCACATTTGCAACCGGAGGAATATACCATTTGCGCAG 158
620 eValThrValIleGlyAsnArgThrPheAspValAspValGluPheA 637
159 CAGGGGGGAGCTTGCCNAGCCGCAACATATCGATTGGGAAACATAC 208
637 la.....ValIleAspSerGluAlaMetGluAsnAspAlaThrTyr 650
209 AAAGCCATCATGTTGGGCCACCTGATGATTCACAGCGCGCTGAAGA 258
650 ..... 650
259 AATATGCGCTACATGTCGCTTTCCGATCGAGGCAACAAATTCATTC 308
651 ..... ValProIleThrThr 656
309 GCCCTTCGACAAACCATGCTCAC...AATCGATTTCGACGAAGCGGTA 355
657 ...ProGluThrThrIleGlnSerGlyLysProThrPheAspLysProAsp 672
356 GTCCCGTTGACGATTCAGCTTTAC... 382
673 ValProLeuAlaAsnAspAlaPheSerValIleAspValTyrAsnLys 689
383 .....GCATCCATGCGACGATACGAAC...ACCATCCGC 416
689 pHeGlyAsnAlaSerValAspAlaAsnThrGlyIleValThrPheThr 706
417 CGAGGCTATGACGGGCCACAG...GCGGCGCTATCCGCTC 457
706 roAlaLysGlyValGlyGlnSerGluProIleThrIleProIle 722
458 CCAAGGCGGAGGATATATACAGCTACGACATAAAGCGGTGCCAA 507
723 ...LysIleValTyrGlnAspGlySerValGlyThrThrAspLeuAlaVal 738
508 AATATCGCGCTACCTGACCGCAACCGCAGCCGCAACGGCTTGC 557
739 ThrValSerLysAsnIleTyrGluAsn...ProGlyGluAsnIlePr 753
558 CGACCGTTCCACAATGCC...GGCGCATGCTCA 589
753 oAlaGlyTyrHisLysValThrPheThrAlaGlyGlnGlyThrSerIleG 770
590 CGCAAGA.....GTAGCGGACGATTCAACGCGCGCAC 624
770 luserGlyThrThrValPheAlaValLysAspLysVal...SerLeuProG 786
625 CGATACAGCCCG...AGCT 641
786 uAspLysLeuProValLeuLysAlaLysAspGlyTyrThrAspAlaLys 803
642 GGACAGATCGGCAATGCCCGCA...GCCCTCAC 675
803 rProGluGluAlaThrGlnProIleLysAlaAspThrGluPheVal 819
676 GGCACATGATATCGTCAAAACATATCGGCGCGCAGAGAAATTC 725
820 SerSerAlaThrLysLeuAspAspIleIleGluAsnProGlyLysAsn 836
726 CGGCGCAGGCGATGCCGTGACGGGTATACGGAAGGCTCAACATTGCTG 775
836 eProAlaGlyTyrHisLysValThrPheThrAlaGlyGlnGlyThrSer 853
776 TCATGACGCGCTTGCTGCTTCCACGCAAAACAAAGATGGCGCGCAT 825
853 lueGlySerGlyThrThrValPheAlaValLysAspGlyValSerLeuPro 869
826 AACGATTTGGCAGATATGCGCAACTCAAGAC...TATCCCGCAGCAGC 872
870 GluAspLysLeuProValLeuLysAlaLysAspGlyTyrThrAspAlaLys 886
873 CATCCGGATTTGGCAGTCCAAACCCCAATGCCGCAAGGATGAAG 922

886 strProGluGluAlaThrGln...ProIleLysAlaAspThrGluP 902
923 CCGTCACCAATATCTTATACGACGCATCCCATCAAGGATGGA... 969
902 heValSerSerAlaThrLysLeuAspAspIleIleGluAsnProGlyAsp 918
970 .....GCTGCCGGGGA 983
919 AsnIleProAlaGlyTyrHisLysValThrPheThrAlaGlyGlnGly 935
984 ATACGCTTGCGGCGCATCACGACATCTCTCAAGCGGTCCCAATGG 1033
935 rSerIleGlnSerGlyThrThrValPheAlaValLys...A 948
1034 GCGGATCGCATGGCGAAGGAATCCGCTC...ACGAC 1074
948 spGlyValSerLeuProGluAspLysLeuProValLeuLysAlaLysAsp 964
1075 AATTTGCGCATGCGCATACGCCAAATACCGCTCCCTTACCATTCGCG 1124
965 GlyTyrThrAsp...AlaLysTrpProGly... 973
1125 AATATCCGTTCAACTTGGAGCAGCTTACGCCAAGAAACATCACCT 1174
974 .....G 974
1175 CCTCACCGTCCGCGCTCAACGCAAAATGCAACTGCGACACCA 1224
974 lAlaThrGlnPro...IleLysAlaAspAspThr 984
1225 CGCCACCGGAAGACAGCGCTTACCGTTGACGATGAAGGTTCCGAATT 1274
985 GluPheValSerSerAlaThrLysLeuAspLysSer...AspAl 999
1275 TGACAGACGCTG...AATATGATACAGCTGATTC 1312
999 aAspLysTyrThrProGluGlnLysValThrThrGluLeuAsnLysG 1016
1313 AAGATATTCGGGGCGGTATACCTAAGCTAAGCTGTG...TTTGAT 1359
1016 lProAspAlaSerGluGlyIleLysAsnLysLysAspLeuProLysAsp 1032
1360 CGCAACCGCATG...GAGGTGATGGAAGCTTATTA 1397
1033 AlaLysTyrThrLysGluLysValAspIleSerThrAlaGlyAsnLys 1049
1398 AFTGACACT...CGTGACGAGGTG 1420
1049 sLysGlyThrValValAlaThrTyrSerAspGlySerSerAspLysValG 1066
1421 AGAAATATGTTACAGAAACGAGAAAGAGCTCAGACTGACTTAA 1470
1066 lValAspValThrValThrAspAsnArgSerAspAlaAspLysTyrGlu 1082
1471 GCCATGCGCAACGAAATGGGAAATTAACAGGTTAGTTTAA 1520
1083 ProThrValGluGlyGluLysValGlu... 1091
1521 TTTTATGAGTGTGATATCATTAAGAAAGCACATACAGAGGAGATA 1570
1092 ...IleGlyLysValAspLeuThrAspAsnValThrAsnLeuProT 1107
1571 GTCTAACCGTGT...GATGTACGGGTG...ATACAA 1602
1107 hLeuProGlnGlyThrThrValThrAspValThrProGlyGlyThrIle 1123
1603 CAACCTGCGCACCTGATTAACATGCGTTATATCAAGGACAGGGAAT 1652
1124 AspThrAsnThrPro...GlyAsnTyrGluGlyValIleGluVal 1137
1653 TAAAGCCTGATGGAAGTTGGAGGTGAAACGAAAGAGTGGGAAG 1702
```





```

::
302 leSerSerThrIleSerValGlnProSerSerSerSerSerProThr 318
989 GCTTGGGCGGCATCAGCGACATCTGTCACAGCGTCCAGATGGGCGC 1038
319 ThrSerThrThrSerValGlnProSerSerSerSerSerAlaProThr 335
1039 ATCCG.....CATGGCCAAAGGAATCCGGCGGTGAGGA 1073
335 rSerAlaThrSerValGlnProSerSerSerSerSerProIleSerS 352
1074 CAATTTGGCGATGCGGCATGCGCAATATCCCTCCCTACATTCCTCC 1123
352 eThrThrIleSerValGlnProSerSerSerSerSerProThrThrSer 368
1124 GAAATTCGCTTCAACTTGAGACAGCGTTCACGCAAGAAACATCACC 1172
369 ThrThrSerValGlnProSerSerSerSerSerAlaProThrThrSer 385
1173 .....TCCACACGCTGGCGCGGTGAGGA 1201
385 aThrSerValGlnProSerSerSerSerSerSerValProThrThrSer 402
1202 AAATGTCAACTGCGACACACCGACCGACGAGCGGTACCGTTT 1251
402 hThrSerValGlnProSerSerSerSerSer.....SerSerThrProThr 416
1252 GACGCTAAAGGCTTCCGATTTTGAGAACGACGTGAATATGATACGA 1301
417 ThrThrSerValGlnProSerSerSerSerSerSerValProThrThrSer 433
1302 GCTCGATATTCAGAAATATTCGGGGCGGTATACCT 1338
433 aThrSerValGlnProSerSerSerSerSerThrPro 445
seq_name: sp_invertebrate:q9yU43

```

```

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mervulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paele J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskay R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003539; AAF49848.1;
DR FlyBase: FBgn0036340; CG11274.
DR InterPro: IPR002483; PWT.
DR InterPro: IPR000230; Ribosomal_S12.
DR Pfam: PF01480; PWT. 1.
DR SMART: SM00311; PWT. 1.
DR PROSITE: PS00055; RIBOSOMAL_S12; UNKNOWN.1.
DR PROSITE: PS00055; RIBOSOMAL_S12; UNKNOWN.1.
SQ SEQUENCE 954 AA: 107592 MW: 572216979110A107 CRC64;

```

```

alignment_scores:
  Quality: 145.00      Length: 646
  Ratio: 0.495         Gaps: 33
Percent Similarity: 45.356      Percent Identity: 22.136

```

alignement\_block:

US-09-303-518D-463 x Q9YU43

Align seg 1/1 to: Q9YU43 from: 1 to: 954

```

66 ACACGCTCAGATTTGGCAACGATCCCTTATCCGCGAGTCTGCACC 115
398 SerArgSerArgArgLeuSerArgSerProMetArgTyrSerArgSerPr 414
116 GTCACGATTTGCAACCGCGGGAATATCCACCTATTCGAGCAGGGGG 165
414 OArgArgPheAsnAsnArgArgArgSerProMetMetAsnPheArgLys 431
166 GAGCTTGGCGCAGCGACGCGCATATCGATTGGGAAACATACAAACCA 215
431 LysLys..... 432
216 TCAGTTGGCCACCTGATGATTAACAAGCGCGCGGTGAAGAAATATCG 265
433 .....ArgGlyLysLysArg. GlyLysLys..... 440
266 GCTAATTTGTCGCGCTTTCCGATCAGCGGCAAAATTCATCCGCCCTTC 315
441 .....HisArgLysMet 444
316 GACACGATGCGTCAATTCGATTCGACGAGCGGCTAGTCCCGTTGA 365
445 TrrGlnHisArgLysLysSerProAsnPheArgLysLys..... 458
366 CGGATTCACGCTTTCCGATTCATTCGAGGAGGATACGACACCATCCG 415
459 .....ArgLysHisTrpLysLys.....ArgHisSerProG 469
416 CCGAGCGGTATGACGCGCACAGCGCGGCGGTATCCGCTCCCAAGGC 465
465 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 485
466 GCGAGGATATATACAGCTACGACATTAAGGCGTTCGCCCAAAATATCG 515

```

```
485 rglnserProglInlnslnArgTyrAsnArgAspHisArglnSerPro 501
516 CCTCA.....CCTGACCGCAAC 535
502 MetGlnAsnGlnProPheArgLysGlnPheSerProHisGlnLysPr 518
536 GCAGCAGCAACAGCGCTGGCGACCGTTTCCAAATGCCGGGGCTATG 585
518 oclYargArgPheSerSer.....ProGlnlnArgTyr.... 529
586 CTGACGCAAGAGTAGCGCAGCATTCACCGCCACCGCATACAGCC 635
530 ..AsnSerArgAspArgTyr..ProAsnSerArgGlnArgArgSerSerPr 545
636 C.....GAGCTGACAGATCGCGCAATGCCCGCAAGCCT 670
545 oclYclYclYclYclYlnHisMetAsnArgTyrAspAsn..ProProAla 561
671 TCACGCGACATGCAGATATCGTCAAAAACATCATCGCGCGCGAGAGAA 720
562 ArgAsn.....ArgArgSerSerSerLysSerSerAla..GlyclYarg 576
721 ATTGCGCGCGAGCGATGCCGTGCAGGGTATACGAGGCTCAACAT 770
576 lnglnArgGlnArgSerAlaSerProProGlnLysArgArgSerSerHis 592
771 TGC.....TGTCATGACGCGCTTGGGTC 793
593 SerArgSerArgSerLeuSerArgAsnArgSerArgSerArgSerSerSe 609
794 TGCCTTCCACCGAAACAGATGGCGGCATCAACGATTTGGCAGATATG 843
609 rGlnHisSerArgLysArgLysPro.....lleclyArgSerS 623
844 GCGCA.....ACTCAAGACTATGCCCGCAGC 869
623 erValGlnTyrAlaGlyProAlaValAsnThrIleAspLeucysArgGln 639
870 AGCCATCCGCGATTGGCGAGTCCAAA..... 896
640 GlnGlnGlnArgLysAsnLysProGlnThrIleIleValLeuAspGlnI 656
897 .CCCCATGCCGCAACAGCAT.....AGAACCGCTCAGCAAT 933
656 yProAlaValArgSerSerArgLysLeuSerLeuArgThrProSerProP 673
934 ATCTTTATGCGAGCCATCCCATCAAGAGGATGGAGCTCCGGGAAA 983
673 heLeu.....LysProHisGlnArg..LeuAlaAla..... 682
984 ATACGCGCTGGGCGGCATCACGGCATCCTGTCAACGCGGTCCAGATG 1033
683 .....AlaLysAlaAlaAlaAla 689
1034 GCGCGATCGCATTCGCGAA.....GGGAATCCCGCGTC 1068
689 lAserAlaAlaValProLysLysSerArgAspGlnSerGlnSerSerSer 705
1069 AGCGCAATTTTGGCGATGCGGCATACGCCAAATATCCCTCCCTTACCA 1118
706 SerAspSerSerAlaGlnSerAspAspSerGln..... 716
1119 TTCCCGAAATATCCGTTCAAACTTGGAGCAGCGTTACGCCAAGAAACA 1168
717 .....GlnGlnProArgArgLysLysGlnP 725
1169 TCACCTCTCAACCGTGCAGCGCAACAGCGCAAAATGTCAACTGGCA 1218
725 rothPasn.....MetGlnThrSerLysGlnLysLysGlnLysValArg 739
1219 GACCAACGCCACCGGAGAGACAGCGTACCGTTGACGGTAAAGGTTTCC 1268
740 ArgSerSerGlnSerSerThrSer..... 747
```

```
1269 GAATTTTGAGAACGACGTGAATATGTATAGCAAGCTCATATTCAGAAAT 1318
748 .....SerGlnGlnSerArgAspGlnAspA 756
1319 TATCGGGGGCGGTATACCTTAAGGCTTAAGCCTGTGTTGATGCGGAACCG 1368
756 snSerSerGly.....AspGlnArgLeu 763
1369 AGATGGAGGTTGTATGAAGCTTAATATA.....TTGACAAAC 1406
764 LysLysLysLeuGlnGlnArgLysSerLysGlnAlaAlaLeuLysGln 780
1407 TCGTAGACAGGTGAGAAAAATGTTCAAGAACGAGAGAGAGTACAGA 1456
780 LysGlnLysGlnHisGlnLysGlnLysAspLysArgArgSerAspLysL 797
1457 GTAGTCACTTAAAGCCCATGCCCAACGAGATGGGAAATTAACACGG 1506
797 ySerLysArgGlyAla..... 802
1507 TTGATTTTAATCATTTTATAGGTGTCATATCAATAAGAACGACAGT 1556
803 .....llecSerSerGlnAspGlnGlnSerThrSerSerArg 814
1557 AACAGAGGCGATAGTCTAACCCGTGTGATGACGGGTGATACAAACA 1606
814 gLysArgGlnHisLysLysThrArgArgAsp...GlnAlaGlyAspAsnS 830
1607 CCTCGGCACCTGATTAACATGGGGTTTATCAAGCGACAGTGGAAATTA 1656
830 erAspSerGlnAspGlnHisValProLysLysLysArgLysLysAlaLys 846
1657 AACCT.....GATGGAAGTTGGAGGTGAAAACGAAAAAGG 1694
847 LysProAlaAspSerSerAspSerAspSerGlnAlaSerSerLysArg... 862
1695 TGGGAAGTGATGACCAACACACCATGTTCCCAAAAGATTGGAGTAGG 1744
863 ...LysLysHisLysLysHisLysLysHisSerLysLysSerLysHis 878
1745 CTAGAAATTAGGCTGAAGTTACTTCGCGTTGGGAAAGTGAATATGCTT 1794
878 lLysLysHisLysArgLysCysSerAlaGlnLysSerArg..... 891
1795 AAGGATATATAATGCGAGGTTACAACT 1821
892 AlaAspSerSerSerGlnLysSerSer 900
```

